BOSTON COLLEGE

THE VAPORFLY EFFECT: INNOVATION OR OMITTED VARIABLES?

A SENIOR THESIS

SUBMITTED TO

THE MORRISSEY COLLEGE OF ARTS AND SCIENCES

DEPARTMENT OF ECONOMICS

BY

WILLIAM R. PETERS

MAY, 2021

Acknowledgements

Thank you to Professor Maxwell, my thesis advisor, my professor, my mentor and my conversation partner. Prior to taking your *Econometrics* course sophomore year, I would have never guessed that I would learn to love running regressions in Stata. Your iterative approach to both econometrics and life inspires me. Our conversations, ranging from running and other sports, to academics and careers, have always brought me joy.

Thank you to Professors Cichello, Grubb and Mortimer for pushing me to pursue my academic interests in Economics to the fullest. Thanks also to Professors Cole and Bannon for making your core classes incredibly enjoyable through your enthusiastic style of instruction and generosity of spirit. Thank you to the Computer Science department for teaching me how to learn independently. And thanks to Tracy Regan for always taking an interest in my life and work and being a great friend.

Thanks to my friends—Erin, Charlie, Liam, Colin, Ruff, and others—for listening to me ramble about a pair of shoes. Thanks for reminding me to both work hard and play hard. To the folks I have gotten to run with over the years—Mike, Dan, Tom, Kamm, Russ, Sean, Aidan, Alex, Quinn and everyone else—thanks for all the conversations and comradery, for keeping me coming back and keeping me humble.

To my family, and, in particular, my parents – thank you for encouraging me to pursue my interests. To my Dad, thank you for inspiring my love of running. And to my Mom, thank you for reminding me that there is so much more to life worth loving.

Chestnut Hill, MA
May, 2021

Table of Contents

Introduction

In sports of human endurance, the role of equipment and technology, and how to regulate it, has long been contested. In the 1980's the advent of carbon fiber bicycle frames shook the world of professional cycling, and the sport chose to accept, rather than outlaw, the arms race of innovation that followed. In swimming, polyurethane and neoprene full-body suits changed the landscape, and the look, of the sport in the 2008 Beijing Olympics; in contrast to the cycling tech, those suits were shortly thereafter banned. In running, the arrival of the Nike Vaporfly 4% shoes in July of 2017 has sparked equal levels of scrutiny. As of now, the shoes that claim to improve marathon performance by 4% have not been banned, and have subsequently sparked an arms race among shoe companies never before seen in the sport.

The question, however, of whether or not the shoes should be allowed in professional marathon contests, remains. While they were initially approved for usage by the International Olympic Committee, the postponement of the 2020 Tokyo Games to the summer of 2021 has re-opened the case. The concern is that, all else equal, these shoes create uneven grounds for competition in a sport that should be won or lost not by technology, but by the fastest person on a given day.

Nike's central claim surrounding these shoes is that the shoes can make athletes "four percent more efficient than Nike's previous fastest marathon shoe," according to their July 17, 2017 press release regarding the shoes (Nike News 2017). Many simplify this quantification of the shoes' effect to mean that the shoes will make a runner 4% faster than they would be in any other shoes, given Nike's position atop the world of running shoe technology.

Figure 1: The Nike Vaporfly 4%'s in the July 2017 Press Release

Source: Nike News

The shoes were designed specifically for distances of half marathon and marathon, and are meant to be worn while racing on roads. In that same 2017 press release, Nike backs up these claims with the key technological features of the shoes: the shoe "pairs a Nike ZoomX midsole (for responsive cushioning) with a full-length carbon plate (intended to minimize energy loss during toe bend without increasing demand for the calf)". The company claims that the carbon fiber plate acts as a spring by improving a runner's foot strike and then by creating leverage to propel runners forward with less effort. Meanwhile, the patented React foam lines the bottom of the shoe to help absorb the force of impact upon landing, thereby reducing the energy lost to impact with the ground. These highly engineered shoes are intended to provide highly engineered results.

Since the release of the shoes in 2017, the arms race among shoe companies to create the greatest and fastest 'super shoe' has been intense. Nike has since released other iterations of the shoe, the latest being the Alphafly NEXT%, and continues to engineer new products that provide additional gains. Competitors, such as Hoka One One, New Balance, and Adidas, have all released shoes that leverage similar technology and make similar claims. While the shoes all leverage the same key piece of technology, a carbon fiber plate in the midsole, each company tries to occupy a niche space in the running world. For example, Hoka One One's shoe targets the marathon and longer, while New Balance's shoe targets short road races, like a 5k. For all intents and purposes, this study treats all three iterations of Nike's shoes (the Vaporfly 4%, the Vaporfly NEXT%, and the Alphafly NEXT%) the same, assuming minimal improvement in outcomes or mechanical variation between the three.

Questions remain, still, surrounding the accuracy of Nike's claim. These shoes may be highly engineered by a leading shoe manufacturer, tested in advanced sports science laboratories, and worn by the fastest professionals, but there is not strong evidence that these shoes actually improve performance by 4%, particularly for everyday runners. Nike's claim has not been verified in the field because, on one hand, running a randomized control trial in the context of marathon running would take a long time and be quite costly. On the other hand, Nike's professional runners are the fastest in the world, having set the recent world records in the marathon and half marathon while wearing the Vaporflys. In short, Nike is not compelled to show true evidence of 4% improvement in everyday runners, and they believe that people will buy into the excitement surrounding the shoes generated by Nike's professional athletes and their performances while wearing the shoes. Thus, the onus is on users and other parties to test Nike's claim and truly determine whether or not these shoes are worth the hype and recognition that they garner. This study sets out to test and evaluate their claim.

The results of this study have implications not only for sport, but also for policy, and even econometrics. Most obviously, the implications of this study are enormous for the entire landscape of running. These shoes are the single largest innovation in the sport since the invention of running spikes to be worn during track races. Since the release, a Nike athlete and Olympic champion, Eliud Kipchoge, became the first man to ever break 2 hours in the marathon, a barrier previously thought of as insurmountable. Along with that historic performance, times have improved for professional marathon runners across the board. The world record in the marathon dropped by 78 seconds from 2:02:51 to 2:01:39, a margin of decline more than double of any world record improvement in the modern era of running. The half marathon record has been improved a handful of times since the release of the shoes, and was most notably broken by four men in the same race at the Valencia Half Marathon in December of 2020. The athletics community wants to know: are these explosive changes happening because athletes are actually getting faster, or do the improvements rely entirely on the advancement in technology?

For economists, the policy and econometric implications are even more meaningful than the athletic implications. The rulings that the World Athletics Association and the International Olympic Committee make with regard to these shoes will address the precedent for technology's role in sport. Officials are forced to rule on the fairness of technology, and thereby must address once again how technology should impact a sport that relies so much on human endurance and performance, in the same way that rubber tracks and shoes with spikes were approved in times of cinder tracks and simple flats. To assess the fairness of using this technology, the officials need to determine if the shoes provide a measureable advantage over other shoes, and then assess whether or not that advantage should be allowed. These rulings are significant for the upcoming 2021 Tokyo Olympics and will set the precedent for other sports and for future technological advancements in running.

The econometric implications of the study are less obvious, but potentially even more interesting to many economists. Below the athletic and political surface of these shoes, lies one of an econometrician's greatest challenges: How does one handle omitted variables and their biases? In determining the actual performance impact of these shoes, the Vaporfly effect, we will be forced to detangle that effect from fitnes*s* effects, which are inherently omitted and difficult to capture. This problem presents space for addressing omitted variable bias in both classical ways, through Instrumental Variables, and creative ways, through difference-in-difference techniques, among others. So, we ask: What extent of impact on the sport can be attributed to individual athletes' fitness levels, and what extent of the impact can be attributed to the super shoes?

Literature Review

Despite the recent release and limited usage of the shoes, there has been a flourish of literature surrounding the shoe technology, as many different groups weigh in on the shoes and their impact. The variety of different contexts of these studies reflects that the scope clearly extends beyond professional endurance sport: economists have studied the effects, legal groups have evaluated the policy outcomes, physiologists have analyzed the energetic costs associated with the shoes, and mainstream publications have presented Nike's claim and subsequent skepticism to the general public.

Popular Press

In 2018, shortly after Nike's release of the first iteration of the shoes, the *New York Times* published their own review of the shoes in their article titled "Nike Says Its $250 Shoe Will Make You Run Much Faster. What If That's Actually True?". The *New York Times* was the first mainstream publisher to put forth both qualitative and quantitative analysis of the shoes out to a mass audience, signaling the broad interest in the topic.

In their analysis, the *New York Times* analyzed more than 500,000 individual race performances between 2014 and 2018 and found that the runners wearing Vaporflys ran, on average, 3-4% faster over the half marathon or marathon distance than categorically similar runners. In their study, they collected the race instances by collecting activity data from Strava, a social media platform on which users, or athletes, can upload exercise activities. Each race instance included a variety of variables that they would later work into models, including duration, distance (half or full marathon), race course and race year specific data, athlete-specific data, such as age and gender, and the shoes that the athlete was wearing. Depending on what athletes recorded in their Strava posts, not all activities, Strava's term for a race instance, included all of these variables; most significantly, "in about one-third of the races on Strava, athletes reported data on the shoes that they wore" (Quealy 3). Nonetheless, with such a large dataset of over 500,000 instances, incomplete data could be handled appropriately for complete analysis.

With this data, the *New York Times* performed their analysis using four models:

1. A statistical model
2. An inter-athlete comparison model for different athletes of similar fitness
3. An intra-athlete model for athletes that switched shoes between races
4. A model that measured the likelihood of an athlete achieving a personal best when wearing the shoes

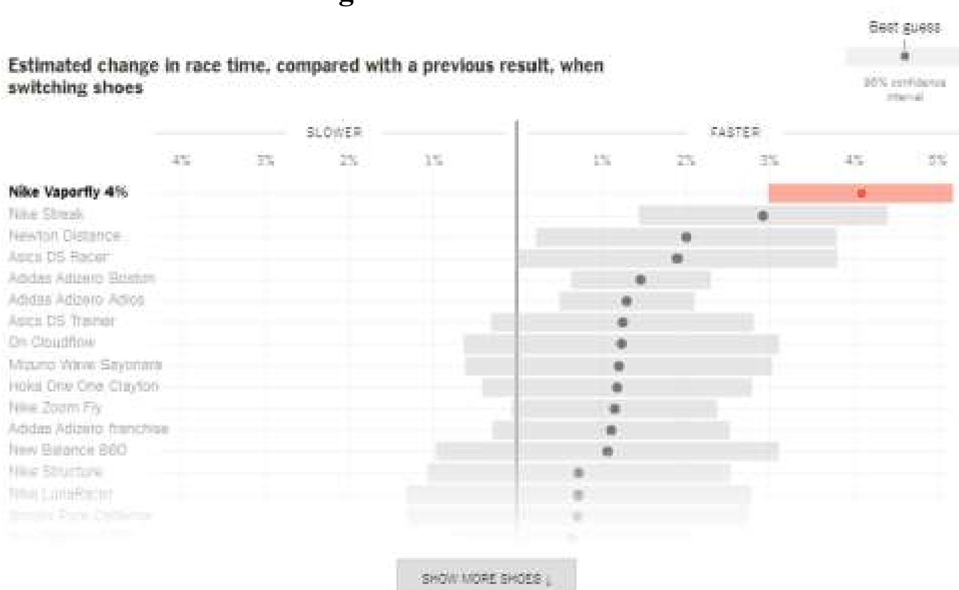
Figure 2: High level results of the NYT's 4 models

Source: Katz & Quealy, NYT

1. Starting with the statistical model, while the *Times* does not specifically lay out their model specifications in the article, they control for athlete-specific and race-

specific factors to estimate how the shoes impact performance: “After controlling for all of these variables, our model estimates that the shoes account for an expected improvement of about 4 percent over a runner’s previous time” (Quealy 4). The authors claim, although do not show, that their results do not change under a variety of model specifications, such as including or excluding athlete training data, using propensity score to determine the likelihood of wearing the shoes, or controlling for the weather on race day. While admitting that this model is potentially biased because it does not use randomized control trial data, Quealy and Katz put together a convincing argument for a Vaporfly effect by comparing the effect of the Vaporflys to the effect of other shoes, as shown graphically in Figure 3, below:

Figure 3: Statistical Model

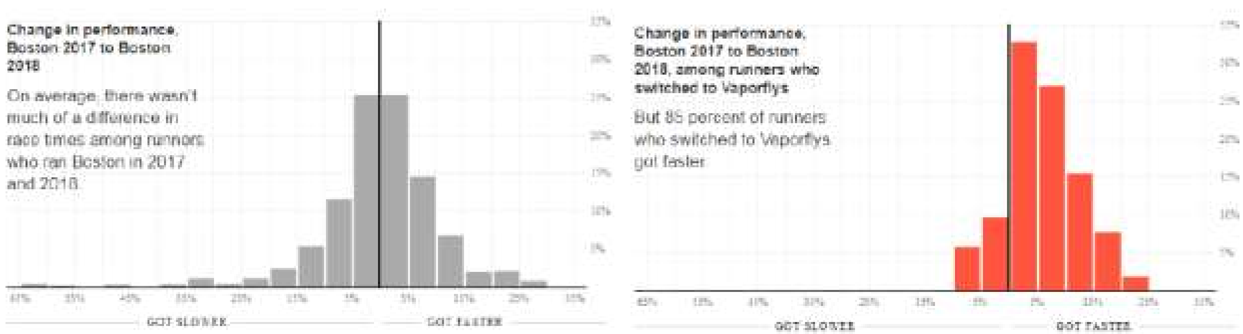


Source: Katz & Quealy, NYT

This figure demonstrates that, while other shoes might also boost performance over a baseline, the Vaporflys are the only shoes that live up to the promise of 4%+ advantage over that baseline.

2. Next, the inter-athlete comparison model makes comparisons between similar athletes over time. The model takes groups of runners that ran the same marathons in the same years, and then looked at how runners that switched to the Vaporflys between races performed as compared to the athletes that did not make the shoe change. This model resembles a difference-in-difference approach, where one group acts as a control group between races, and another receives a treatment (ie, switching to the Vaporflys) between races. When looking at runners that ran the Boston marathon in both 2017 and 2018, those that switched to the Vaporflys in 2018 generally performed better than the rest of the group, as demonstrated by this side by side graphical comparison:

Figure 4: Inter-athlete Comparison Results



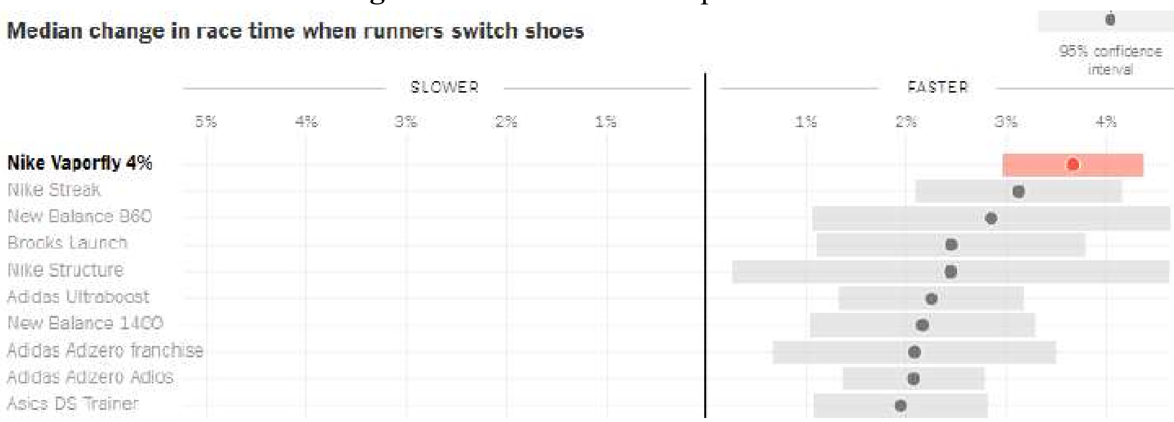
Source: Katz & Quealy, NYT

While the group of runners that made the switch in shoes between years is small, only 52 of 1,275 runners that ran Boston both years, the shift is dramatic.

3. The intra-athlete model aggregated the effects of a switch to Vaporflys within individual athletes over time. This model took athletes who had uploaded at least 5 marathons to Strava in the 2014-2018 time period and then tried to estimate how the switch to the Vaporflys affected their performance. By focusing on athletes with many races uploaded, this model attempts to more accurately capture the

natural improvements that most athletes see as a result of consistent training over time. Then, the model uses the race in which the athlete switches to the Vaporflys to estimate the shoes' effect. The graphic below summarizes the results of not only the effect of a switch to Vaporflys, but also to other shoes:

Figure 5: Intra-athlete Comparison Results

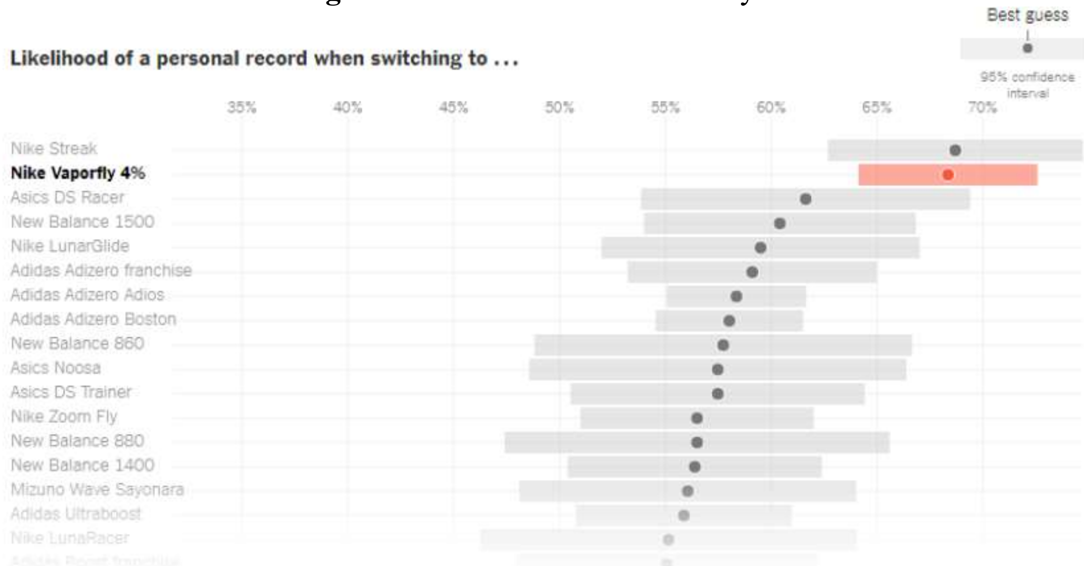


Source: Katz & Quealy, NYT

While this model focuses less on the effects of factors like course and weather, it still makes a convincing case for the improvement that the shoes provide, but at a magnitude lower than 4%.

4. The final model seeks to display how likely runners were to run their fastest time, or personal best (PB), when they made the switch to the Vaporflys, as compared to other shoes. While the authors do not specify the other variables that went into this model, such as race day conditions or training leading up to a PB-performance, the focus is on whether or not the Vaporflys were worn during these special performances. This model again focuses on how runners fared when the switched the Vaporflys in comparison to a switch to other shoes. Yet again, the results are convincing, with nearly 70% of the runners that switched setting a PB, as shown below in Figure 6:

Figure 6: Personal Best Probability Model



Source: Katz & Quealy, NYT

Of interest with these results, however, is the fact that a different shoe, the Nike Streak, actually accounted for a slightly higher probability of running a personal best when switching to those shoes. It is possible that this trend reflects the fact that the Nike Streak is a more accessible, and thus more widely used, shoe, whereas the Vaporflys' $250 price tag puts them out of reach for many. Nonetheless, the case for the Vaporflys is strengthened.

As the *New York Times*' varied analyses suggest, the Vaporflys seem to have some effect. However, by their own admission, there are still plenty of shortcomings to the analysis that make the actual magnitude of the effect under suspect. The analysis failed to control for many potential forms of bias:

- The most significant omission is likely their lack of quantification of any fitness effect, or the effect of a change in fitness between races. While there are methods in place that might handle that fitness effect in the inter-athlete and intra-athlete models, Quealy and Katz do not outright address the existence of this omitted variable in any of their analysis.

- Other missed biases include self-selection bias and reporting bias, where athletes that are poised to run fast decided to wear the shoes, and athletes that ran well decided to report their performance, respectively.

The authors openly admit that their models are susceptible to these types of biases, making their estimates less certain. Despite these holes in their analysis, the *New York Times* still presents a compelling case for a Vaporfly effect existing in some form.

An Econometric Analysis

In "An Observational Study of the Effects of Nike Vaporfly on Marathon Performance," Guiness et. al. took a narrower, but more sophisticated, approach to analyze the Vaporfly effect. Instead of focusing on a broad spectrum of runners, this study from Cornell University systematically selected professional and semi-professional athletes' performances from 2015-2019, a window spanning evenly from the two years before the release of the shoes and two years after the release of the shoes. The study referenced Quealy and Katz's work with the *New York Times*, and intentionally focused on elite and sub-elite marathoners instead of the general running population for expressed reasons:

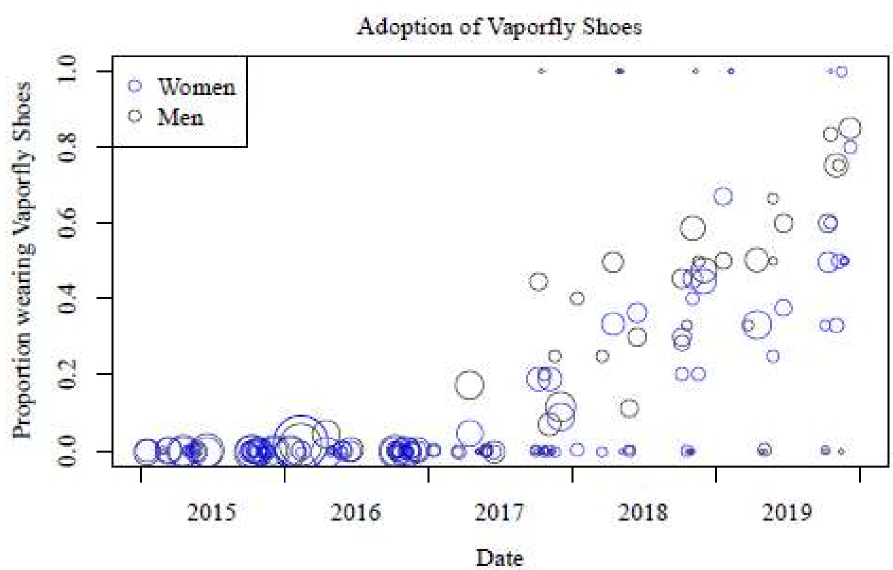
- They wanted to focus on the subsection of runners for whom the impending regulations had the greatest impact.

- They wanted to focus on runners who had proven themselves to have run fast marathons, defined by a minimum time cutoff, before wearing the Vaporflys in an attempt to control more effectively for fitness of athlete-specific effects.
- They avoided some selection bias by collecting "an exhaustive sample of athletes who met a minimum performance standard at one of 22 of the largest marathon venues in 2015 and 2016," instead of the self-reported Strava data used by Katz and Quealy (Guinness 2).

In controlling for athlete ability and race conditions, the group found 90% confidence intervals of 1.4%-2.8% improvement in times for male athletes and 0.6%-2.2% improvement in times for females.

The study collected race data from major US and Canadian marathons in 2015 and 2016 to capture the pre-Vaporfly results, and results from marathons in 2017, 2018 and 2019 for results that included Vaporflys (see Appendix). Adoption was not immediate, nor was it universal, as one would expect with any new technology.

Figure 7: Adoption of Vaporflys over Dataset Timeframe



Source: Guinness et. al.

The above graph demonstrates that adoption was low starting in 2017, and gradually rose over the course of the subsequent years. Each circle represents a race, with the size of the circle corresponding to the number of athletes that met researchers' selection criteria in that race. The vertical position of the circle shows the proportion of runners in the given race wearing the Vaporflys. The plot demonstrates steady adoption of the shoes, starting low and passing an average of 50% adoption of the shoes by 2019. In this study, men and women were split into different groups for the purpose of analysis, so as to account for any differences in how the shoes impacted the sexes physiologically differently. These physiological differences can be attributed to mechanical differences in the typical strides of male and female runners, as well as height and weight differences.

In their analysis, Guinness et. al. used a relatively simple fixed effect statistical model of the performances, as follows:

Figure 8: Guinness Model Specifications

$$y_i = \text{marathon time in minutes for performance } i$$

$$x_i = \begin{cases} 1 & \text{if Vaporfly shoes worn in performance } i \\ 0 & \text{if Vaporfly shoes not worn in performance } i \end{cases}$$

$$j(i) = \text{label for athlete who completed performance } i$$

$$k(i) = \text{label for marathon course associated with performance } i$$

$$\ell(i) = \text{label for individual marathon race associated with performance } i$$

$$\text{Untransformed:} \quad Y_i = b_0 + b_1 x_i + U_{j(i)} + V_{k(i)} + W_{\ell(i)} + Z_i$$

$$\text{Log Untransformed:} \quad \log Y_i = b_0 + b_1 x_i + U_{j(i)} + V_{k(i)} + W_{\ell(i)} + Z_i$$

Terms		Assumptions	Description
b_0, b_1		non-random parameters	b_1 = Vaporfly effect
$U_1, \ldots, U_A$	$\overset{ind}{\sim}$	$N(0, \sigma_1^2)$	runner effects
$V_1, \ldots, V_C$	$\overset{ind}{\sim}$	$N(0, \sigma_2^2)$	course effects
$W_1, \ldots, W_R$	$\overset{ind}{\sim}$	$N(0, \sigma_3^2)$	individual race effects
$Z_1, \ldots, Z_n$	$\overset{ind}{\sim}$	$N(0, \sigma_4^2)$	residual effects

Source: Guinness et. al.

In short, the parameter of interest, or the Vaporfly effect, is measured by b_l, the coefficient of the dummy variable, x_i, which is a Boolean representation of whether or not Vaporflys are worn. Each runner has its own offset term, $U_{j(i)}$, to account for varying athlete abilities, each course has an offset term, $V_{k(i)}$, to account for differences in course variation, each individual race has an offset term, $W_{l(i)}$, to account for differences between the same races from year to year, and each instance in the dataset has a term, Z_i, to account for any other residual effects. They used R's lmer package, or the linear mixed effects, to fit their model to the given parameters. The results were promising, showing significant time reduction effects from the use of Vaporflys:

Figure 9: Guinnes et. al. Results Table

	men minutes	women minutes	men log minutes	women log minutes
	estimate (s.e.)	estimate (s.e.)	estimate (s.e.)	estimate (s.e.)
b_0	139.69 (0.59)	159.83 (0.81)	4.94 (0.004)	5.070 (0.0050)
b_1	**−2.95 (0.60)**	**−2.18 (0.81)**	**−0.0209 (0.0041)**	**−0.0135 (0.0049)**
σ_1	4.175	6.40	0.030	0.041
σ_2	1.852	2.33	0.013	0.014
σ_3	1.874	2.43	0.013	0.015
σ_4	4.108	5.02	0.028	0.030

Source: Guinness et. al.

As the output shows, wearing the Vaporflys was shown to be associated with almost a 3-minute improvement for men, and just over 2-minute improvement for women, a 2.09%-time reduction and 1.35%-time reduction, respectively. While these estimates do not match the 4% claim from Nike, they still promise a significant advantage.

In his analysis, Guinness opens himself up to critiques in his approach:

- The biggest critiques come from the assumptions that were made in building the statistical model. One major assumption is that the shoes affect all runners of the same sex equally. Two key threats to this assumption would be, first, that this does not account for biomechanical differences between athletes of the same sex, and, second, that the newer iterations of the shoes, while not distinguished in the study, could also affect runners differently.

- Additionally, the model assumes that, in any given race, times run follow a normal distribution, when in reality it is much "more likely to run 5 minutes slower than expected rather than 5 minutes faster; when things go wrong in a marathon, they can go really wrong" (Guinness 8).
- Finally, other minor concerns include that this model does not account for runners who did not finish their races, along with the fact that the shoes had to be manually identified in pictures, leaving room for human error.

Otherwise, this study puts forth another convincing argument in favor of a Vaporfly effect of 1.4%-2.1%.

A Physiological Analysis

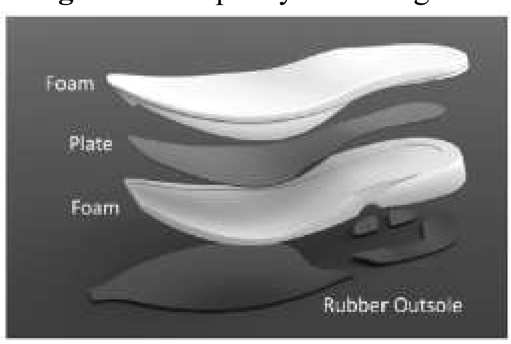
In the discussion of the physiological reasoning behind the improvements, Guinness et. al.'s paper references Hoogakamer et. al.'s 2018 laboratory study, "A Comparison of the Energetic Cost of Running in Marathon Racing Shoes." That study focuses more specifically on the physiological effects of the shoes by studying 18 "high caliber athletes [running] six 5-min trials" in a laboratory environment (Hoogakamer 1009). The trial calculates the Vaporfly effect in the form of change in energetic cost, measured in watts per kilogram, when an athlete switches from two other premier marathon shoes to the Vaporflys. This study found that the Vaporflys "lowered energetic cost of running by 4% on average" (Hoogakamer 1009). While a reduction in energetic costs is not exactly translatable to a time reduction in a race, the estimate sticks as Nike's metric of improvement due to the shoes. The most notable outcome from this study was how strongly it endorsed Nike's claim of a 4% improvement; the paper goes further to predict that an athlete will soon achieve the sub-2-hour marathon while wearing the shoes, which happened shortly after conclusion of the study. It is worth noting in evaluating these findings that this University of Colorado study was funded by Nike, clearly reflecting their desire to prove the shoes' effectiveness to runners and scientists alike.

The paper starts by asserting that three physiological parameters generally decide how long a certain velocity, or pace, can be sustained while running: runner's maximal rate of oxygen uptake, their lactate threshold, and their energetic cost of running. The primary assumption of the study is that runners of similar levels, particularly at the professional or elite level, have similar lactate threshold levels and oxygen uptake rates. Thus, running economy, or efficiency, which then determines how long a certain pace can be held for that athlete, can be expressed by way of energetic cost at that certain pace, measured in Watts per Kilogram (W/Kg). Following the simple math in this three-part equation of lactate threshold, oxygen uptake and energetic costs, low energetic costs athletes outperform high energetic cost athletes when the other two factors are assumed to be effectively the same. Thus, "if an athlete can lower their energetic cost to run at a specified velocity, then they should be able to run faster with their existing physiological capacities" (Hoogakamer 1010). As such, the study seeks to evaluate whether or not Nike's shoes lower energetic costs more than their alternatives.

A variety of factors influence the energetic cost associated with certain pairs of shoes, including shoe weight, shoe cushion and shoe stiffness or springiness. All of these factors are logical influencers of energetic costs; lighter shoes require less energy to move, more cushioned shoes reduce the energetic cost placed on the feet and lower legs when striking the ground, and springier shoes reduce the cost of physically moving forward. Shoe cushioning has been a major focus of shoe manufacturers for years. The goal of such focus is to create lightweight foam to be placed in the midsole that is both compliant, in that it reduces the energetic cost of impact, and resilient, in that it can store and return mechanical energy associated with foot strike. Shoe companies have made large improvements in this area over the last 20 years in their development of lightweight, highly resilient midsole foam, recently "shown to reduce the energetic cost of running by ~1%" (Hoogakamer 1010). These improvements, while not as striking as a 4-5% improvement, are significant.

The new breed of 'super shoes' goes a step further in its attempt to enhance the mechanical energy saved by running shoes by adding a plate in between layers of foam within the midsole, as shown in Figure 10.

Figure 10: Vaporfly Sole Diagram



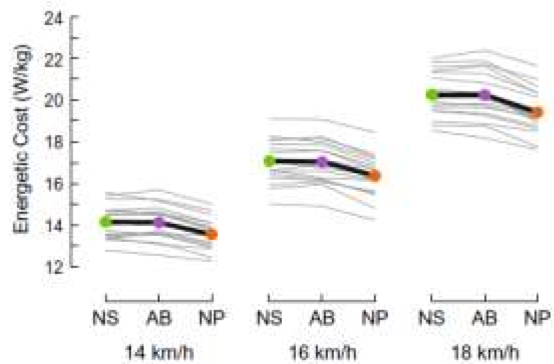
Source: Hoogakamer et. al.

This midsole carbon-fiber plate acts as a spring, helping runners propel into their next stride following contact with the ground. These plates have also been shown to reduce the energetic cost of running by around 1% according to the researchers in this study. Thus, combined with resilient foam and incredibly light upper mesh covering the top of the shoe, Nike has created shoes that they believe significantly decrease the energetic costs of running.

To test that belief, Hoogakamer et. al. tested in a lab the Vaporflys against Nike's Zoom Streak 6, formerly their standard marathon racing shoe, and the Adidas Adizero Adios BOOST 2, the shoes worn by Dennis Kimmetto in his September 2014 world record performance. For reference, the *New York Times* study found that runners were more likely to set a personal best in the Nike Zoom Streak than any other running shoe, including the Vaporflys. Prior to any athlete being tested in the shoes, the shoes underwent mechanical tests by way of lab technology that could measure mechanical energy storage and return. Combining the results of compliance and resilience tests, the Vaporflys were found to return twice the amount of mechanical energy than either of the other shoes, a huge improvement. Following the mechanical tests, tests on human subjects were performed. To make the tests as even as possible, 18 high-caliber, male runners all with very similar physical features and running performance were selected for analysis. The runners visited the lab for testing a total of 4 times each, each time running six 5-minute segments, two at 14 km/h, two at 16 km/h, and two at 18 km/h. The first lab test set a baseline in their own shoes, and then the subsequent three lab visits featured one of the three test shoes.

By controlling for a variety of different variables and recording others, such as the lactate threshold during the 5-minute repetitions, the scientists were able to measure the average energetic cost associated with running in each shoe through a series of advanced sensors and tests placed on the treadmill and on the athlete during the experiment. Results are graphically depicted in Figure 11.

Figure 11: Energetic Cost Differences



Source: Hoogakamer et. al.

At each speed, the Vaporflys (NP) demonstrate energetic cost lower than the other two pairs of shoes, the Nike Streaks (NS) and Adidas Boosts (AB). These differences reflect an average energetic cost 4.01% lower than the other two shoes, thus validating Nike's 4% claim. In the context of the marathon, that 4.01% reduction in energetic cost "should translate to ~3.4% improvement in running velocity at marathon world record pace … and not since 1952 has the men's marathon record been broken by more than 3.4% in one race" (Hoogakamer 1016). As a result, this study clearly demonstrates how revolutionary these shoes are for the sport.

Regulatory Analysis

Finally, two further pieces of literature give a snapshot of the regulatory implications of the results found in the aforementioned studies as well as the proposed work of this study. First, in Dyer's "A Pragmatic Approach to Resolving Technological Unfairness: the Case of Nike's Vaporfly and Alphafly Running Footwear" and then in

O'Grady & Gracey's "An Evaluation of the Decision by World Athletics on Whether or Not to Ban the Nike Vapor Fly Racing Shoe in 2020," medical and legal scholars, respectively, give insight as to why the World Athletics Association continues to allow this new technology in professional running, and propose practical ways of formulating future rulings. In short, both articles recognize that there is mounting evidence that the Vaporfly effect does exist to some degree. However, as information currently stands, there is not enough evidence to suggest that these shoes provide an unfair advantage or encourage coercive behavior on behalf of athletes' sponsors.

Dyer's analysis of the Vaporfly effect takes a look at the question of unfairness as a result of the shoes' effect. In his analysis, Dyer takes a pragmatic approach, which, in the context of sport, attempts to evaluate whether or not the use of the technology is in the best interest of the sports' future. Figure 12 lists the evaluation criteria:

Figure 12: Pragmatic Analysis Criteria

Criterion
Harm or health (to the athlete or others)
Unnaturalness
Unfair advantage
Coercion
Safety and spectator appeal
Integrity of the game, harm to or advantage over the sport itself, or the 'spirit of the sport'
Deskilling and reskilling
Dehumanisation
Cost (or excess cost)
The internal goods of a sport
Equal opportunity or access

Source: Dyer et. al.

These criteria provided the basis for the study's core research questions, the outcomes to which would be suitable for sports stakeholders to act upon. Using those criteria, the questions that Dyer goes on to pose and answer pragmatically took form similar to the following: “Are the Nike Vaporfly/Alphafly shoes harmful to the health of the athlete using them?... Does the use of the Nike Vaporfly/Alphafly shoes affect the integrity of the sport or provide an advantage over the sport itself?... Are the Nike Vaporfly/Alphafly shoes inaccessible to some athletes?” (Dyer 3). Some of these questions, such as the questions relating to the criterion of unnaturalness or athlete harm require only brief explanations: all shoes would be considered unnatural, but they have been widely accepted in the sport for over 100 years, and these shoes show no evidence of athlete harm. These outcomes do not object to the continued use of this shoe technology. However, the questions relating to the unfairness of the technology and the equal access to it required much lengthier explanations.

Dyer does not question whether or not the shoes provide a mechanical advantage in comparison to other shoes. Referencing Hoogkamer’s paper, Dyer notes that the shoes are shown to have clear reductions in energetic costs, which would undoubtedly contribute to an advantage. The existence of an advantage is deemed passable because of the wide variation of shoes and technology that already exist and are allowed in the sport. Dyer does not justify this variation in terms of any statistical measure of outcome variance. Thus, unfairness of the technology is determined not by the existence of an advantage, but the magnitude thereof. In order to assess the fairness of that magnitude from a pragmatic perspective, Dyer relies on the performance improvement index (PII), a measure “developed to demonstrate unusual leaps in performance that could then retrospectively be attributed to technological change” (Dyer 4). The PII is calculated with a relatively simple formula: $[(t1/t2)^2 - 1] * 100$, where $t1$ is a first performance divided by $t2$, a second performance, in order to demonstrate the magnitude of change between the two. Dyer concludes his argument by stating that there is a common misconception that the world records in the marathon and half marathon made unusual jumps with the advent of these shoes, and backs up his claim with the following table summarizing PII changes:

Figure 13: PII Changes

	PII positive change range	PII largest change since 2016	World record PII change ranking
Marathon: men's	0–7.2	2.1	14th/38
Half marathon: men's	0–7.5	0.98	11th/32
Marathon: women's	0.1–9.2	2.0	21st/30
Half marathon: women's	0–6.2	0.7	15th/27

Source: Dyer et. al.

In short, Dyer uses these measures to argue that the record-breaking performances in these shoes did not present any abnormal changes in performance, and improvements in footwear and running surfaces have been happening all along. Thus, Dyer claims that the Vaporfly effect does not present an unusual outcome, and that the playing field of running technology has never truly been level. These are strong claims for shoes that would produce the first sub-2-hour marathon and a slew of world records almost instantly after being released.

Many of the other criteria, such as coercion, reskilling, dehumanization, cost and 'spirit of the sport,' are quickly moved through, as the Vaporflys are easily deemed passable just as any other new pair of shoes would. The next key dilemma comes with the criterion of accessibility. Because many pros are sponsored by specific brands, the shoes available to them are often limited. As a result, an athlete sponsored by New Balance or Hoka One One would not be able to wear Nike's Vaporfly shoes. Dyer references cases in other sports, such as swimming and cycling, that faced similar accessibility challenges posed by athletes' sponsors. Some of these cases ended in bans, while others not, depending on the scenario. The belief that Nike's creation of this shoe will result in other brands matching the product with their own versions of a 'super shoe' provides evidence against a ban, while the enforcement of intellectual property rights or the inability of other brands to compete would encourage the ban. However, while the Nike-specific technology will not be available to other brands' athletes, the core principles of the technology can certainly be implemented by other brands, thereby debunking this concern of accessibility. Ultimately, the framework used to evaluate the Vaporflys finds seven points in favor of continued use of the technology and three points of contention. As a result, Dyer's primary recommendation, from a pragmatic perspective, is to continue to allow the shoes to be worn, but with continued vigilance and willingness to revisit the regulations. Mainstream media companies such as NPR and ABC have turned to Dyer for opinions on the shoes, yet his ultimate judgement avoids setting a firm stance.

Because of the mounting evidence that the Vaporflys did indeed have some effect on racing outcomes, the World Athletics Association (WAA) had been drafting some legislation since October of 2019 to address any unfairness that the shoes may cause in the sport. In O'Grady and Gracey's "An Evaluation of the Decision by World Athletics on Whether or Not to Ban the Nike Vapor Fly Racing Shoe in 2020," the authors take a look at how the WAA went about that decision making process. The delay between the release of the shoes in 2017 and the decision from the WAA is notable, as races needed to be run and studies need to be conducted for the WAA to determine whether the shoes actually made a difference in performance. After Eliud Kipchoge's 2019 1:59:41 marathon performance, among other stellar performances by athletes wearing the shoes, and multiple studies produced suggesting the shoes truly did create a near 4% advantage in competition, the WAA was led to begin the evaluation process. Kipchoge set the official marathon world record in the shoes in September 2018, running 2:01:39 at Berlin, and Brigid Kosgei took down Paula Radcliffe's longstanding world record by running 2:14:04 in Vaporflys the 2019 Chicago Marathon. The women's half marathon world record has been crushed twice by athletes wearing the shoes since their release: first by Ababel Yeshaneh in February 2020, running 1:04:31, and then by Ruth Chepngetich in April 2021, running 1:04:01. The studies by teams like Hoogakamer et. al. from the University of Colorado as well as the *New York Times* team have only further confirmed the power of the shoes in the public forum. Then, that evaluation process could not simply consist of handing down a binary banning ruling on the shoes, but had to be specific about which features and specifications of the shoes would not be permitted in the sport. Thus, this proceeding was no simple process.

The WAA's evaluation of the legality of the Vaporflys in sport was modularized into multiple parts. First, the WAA sought to determine if the shoes provided a mechanical advantage over other shoes that would be fundamentally unfair. Gracey and O'Grady compare this analysis to the cases of the Speedo LZR Racer swimsuit, which was ultimately banned by the swimming governing body, and the case of the clap skates worn by speed skaters, which was ultimately banned at the short track level but not at the long track level. The key distinction between these two rulings had to do with how the mechanics of the technology affected the racer. In short, in cases where the technology affected the external racing environment, a ban was deemed necessary, while cases in which the technology only affected the athlete wearing it did not receive a ban. Thus, the swimsuit was banned while the skates were not.

The WAA considered the case of the Vaporflys in this context: Do the Vaporflys change the race, or change the racer? As discussed in Hoogkamer's paper, the shoes clearly do change the biomechanics and energetic costs of an individual racer, but do not manipulate the external environment that the athlete is participating in. So, while the WAA recognizes that a mechanical advantage is present in the Vaporflys, they do not find that advantage worthy of banning, as other shoe companies are able to compete with this technology. Nonetheless, in order to address the mechanical advantage, the WAA did modify Rule 5 of Book C in their Constitution and Book of Rules to limit the effect that the mechanical advantage would provide. The rule was modified as follows: "'the sole must be no thicker than 40 mm' and 'the shoe must not contain more than one rigid embedded plate or blade (of any material) that runs either the full length or only part of the length of the shoe. The plate may be in more than one part but those parts must be located sequentially in one plane (not stacked or in parallel) and must not overlap'" (O'Grady 3). The Vaporflys very narrowly meet these criteria, and are thus deemed passable. This modification sought to make more clear to shoe manufacturers the limits of the innovation that they could make in their technology moving forward.

Next, the WAA's evaluation of the shoes focused on the universality of athletics. The basic premise of this universality is that the sport, and all advantages within it, should be reasonably accessible not only to all professional athletes, but all amateurs as well. That is to say that they should be available for purchase to all people, and should be reasonably accessible. While the price tag of $250 makes the shoes some of the most expensive in the sport, it was not considered inaccessible by the WAA. This question of accessibility also brings attention to the contracts that athletes have with different sponsors that may hinder their access to the shoes. However, because it is believed that athletes sign these contracts at will, the WAA does not find these shoes inaccessible to athletes sponsored by shoe brands other than Nike because these athletes could theoretically switch to a Nike sponsorship or find sponsorship from other non-shoe brands. In order to definitively modify the ruling on accessibility, the WAA ruled to require any shoes worn in competition must be available for sale to the general public for at least four months prior to the competition. This marks a large shift in policy from the WAA, as athletes will no longer be able to compete in their sponsors' prototype shoes. This ruling, however, was deemed sufficient by the WAA with regard to the universality of athletics. Now, one can expect to see both professionals and amateurs wearing the same shoes at the start of a marathon.

O'Grady and Gracey's recommendation to the World Athletics Association and other regulators is as follows moving forward: Keep a vigilant eye on the advancement of this technology, and reassess the ruling as often necessary. There are early signs that shoe companies, like Nike, are going to push the regulations passed down to the absolute limit, requiring vigilance on behalf of the governing body. While specific rules were modified, surely Nike and others will continue to innovate with other pieces of technology in the continued arms race of running shoe technology.

With an improved understanding of the mechanics of the shoes as well as the controversy and rulings surrounding them, it seems that the Vaporflys have generated plenty of hype, if nothing else. Guinness et. al. and the *New York Times*' team provide inspiration for how the shoes' effect can be measured in the wild. Analysis of the effect they produce in actual races will be most telling.

Data & Methodology

Introduction

Analysis of real world race results is necessary to empirically evaluate Nike's claim on a practical level. While Nike may have proven their claim mechanically in labs, testing the strength and responsiveness of their technology, those tests are irrelevant if the technology's impact does not show up in real world data. To conduct this analysis, a two-pronged approach will be taken:

- First, Guinness et. al.'s "Observational Study of the Effects of Nike Vaporfly on Marathon Performance" is revisited. The goal of this first section is replication and, if possible, improvement. By replicating the study, we will have created a basis for further analysis and comparison; further, by including more recent data in the dataset used for Guinness's study, the results of the study will be strengthened.
- For the second analysis, publicly available Strava marathon data is collected and analyzed in a similar fashion to the *New York Times'* analysis. While Guinness' study focuses on professional runners, this second study hones in on the Vaporfly effect in the context of everyday amateurs.

Guinness Replication

Included in Guinness et. al.'s paper is a link to Joe Guinness' GitHub repository, or online storage location for program scripts and files, which details the specific steps necessary to replicate the study. This convenient resource allows for speedy replication of the study as well as clear opportunities for improvement on the study, most easily by way of expanding the dataset. The following work is grounded largely in the directions found in that repository, along with the analysis discussed in the paper, "An Observational Study of the Effects of Nike Vaporfly on Marathon Performance."

Data Needs

As discussed in the literature review, the Guinness et. al. study takes a relatively simple approach in collecting and analyzing marathon data for their study. In short, Guinness et. al. takes a two-pronged approach to collect all of their data:

- First, they use a Python script to scrape a website, marathonguide.com, for results from certain marathons that would qualify as 'elite,' or faster than 2:24 for men and faster than 2:45 for women. Prior to scraping the data, they compile a list of 109 marathons that were reasonably competitive to collect their data from (see Appendix for full list of marathons).
- Then, they analyze whether or not a runner was wearing the Vaporflys in a given race by manually inspecting race photos from each race and indicating whether or not the runner had the shoes on.

After collecting the data as so, Guinness prepares the data for analysis through a series of cleaning steps, and then runs relatively simple regression analyses on the data. For their model specifications, they split the dataset to study men and women separately, and then run untransformed and log-transformed model specifications. They use a linear mixed model with fixed effects for the race, year, and individual athletes. Their findings are as follows in Figure 9.

Figure 9: Guinnes et. al. Results Table

	men minutes	women minutes	men log minutes	women log minutes
	estimate (s.e.)	estimate (s.e.)	estimate (s.e.)	estimate (s.e.)
b_0	139.69 (0.59)	159.83 (0.81)	4.94 (0.004)	5.070 (0.0050)
b_1	**−2.95 (0.60)**	**−2.18 (0.81)**	**−0.0209 (0.0041)**	**−0.0135 (0.0049)**
σ_1	4.175	6.40	0.030	0.041
σ_2	1.852	2.33	0.013	0.014
σ_3	1.874	2.43	0.013	0.015
σ_4	4.108	5.02	0.028	0.030

Source: Guinness et. al.

So, while not as explosive as Nike's 4% claim, Guinness et. al. find that the Vaporflys provided a ~2.1% improvement for men, and a ~1.4% improvement for women.

In order to replicate this work, the first requirement is to simply fork, or copy, Joe Guinness' repository. This repository has some data already collected and stored in it, eliminating the need to repeat certain parts of the data collection process. For one, the data corresponding to the manual photo inspection is already completed and stored in the repository, saving the need to look through thousands of race photos. Results from one 2020 marathon, the Marathon Project, will be added to the dataset, so some of this manual data collection is necessary, but minimally so. Additionally, the repository comes pre-loaded with data for assigning athletes unique identifiers, and a file containing matches for any misspellings of athlete's names. For the remainder of the data necessary to replicate the analysis, the repository includes a file, reproduce.txt, with instructions on how to go about collecting and analyzing the data.

In order to understand the methods of data collection, first one must understand the data requirements for analysis. The dataset used for analysis is ultimately relatively simple; one instance of the dataset is composed of a result, or time for a given athlete's performance at a race, the race at which the athlete competed and its year, the athlete's name and unique identifier, the athlete's gender, and a variable capturing whether or not the athlete was wearing Vaporflys during the performance of not, given by a simple true or false. The majority of this data likely, and appropriately, sounds like a typical race results page that one might find at any given race: a table containing athlete name, time, gender, place and age. In order to collect that portion of the dataset, Guinness' repository utilizes a Python script to scrape the data from a website that houses many such results pages, marathonguide.com, pictured as follows:

Figure 14: Marathon Results Site

2020 The Marathon Project, Overall Results 1

Have you run this race?
Then tell us about it.

More Results:

Last Name, First Name (Sex/Age)	Time	OverAll Place	Sex Place	BQ*
Martin Hehir (M)	2:08:59	1	1	BQ
Noah Droddy (M)	2:09:09	2	2	BQ
Colin Bennie (M)	2:09:38	3	3	BQ
Scott Fauble (M)	2:09:42	4	4	BQ
Ian Butler (M)	2:09:45	5	5	BQ
Scott Smith (M)	2:09:46	6	6	BQ
Mick Iacofano (M)	2:09:55	7	7	BQ
Benjamin Preisner (M)	2:10:17	8	8	BQ
Nathan Martin (M)	2:11:05	9	9	BQ
CJ Albertson (M)	2:11:18	10	10	BQ
Emmanuel Roudolff (M)	2:11:20	11	11	BQ
Colin Mickow (M)	2:11:22	12	12	BQ
Reid Buchanan (M)	2:11:38	13	13	BQ
Kevin Lewis (M)	2:12:02	14	14	BQ
Cam Levins (M)	2:12:15	15	15	BQ

Source: marathonguide.com

The Python script takes advantage of two key packages, requests and BeautifulSoup, to scrape pages such as the one pictured above and stored them in clean CSV files for later analysis. The requests package automates the visiting of the 111 race results pages and grabs the HTML, and then BeautifulSoup package manipulates the HTML to collect the variables of interest on the page (all columns in Figure 14). Once collected, this data is stored to be cleaned and analyzed later on. This scraping provides all of the athlete data besides whether or not the athlete was wearing Vaporflys in the race of interest. While that data is largely already collected by Guinness et. al., 2020 marathon results were manually collected. Coming from a variety of websites, race photos were similar to the below:

Figure 15: Photos from the Marathon Project, December 2020

Source: runnerspace.com photo gallery

After collecting all of the data in these two different formats, the data needs to be cleaned before analysis could begin. First, performances that are missing any of the data necessary for analysis are dropped from the dataset. Additionally, race times that do not meet the ‘elite’ criteria of qualifying for the US Olympic Team trials (2:45 for women, 2:18 for men) are also dropped. The dataset scraped of race results from marathonguide.com is merged with the pre-existing dataset containing unique athlete identifiers, so that any name misspellings are avoided, and then finally that dataset is merged with the dataset of manually inspected photos based on the unique athlete identifiers and the race date to correctly indicate whether each individual performance was by an athlete wearing the Vaporflys or not. Figure 16 presents an extract from the final dataset:

Figure 16: Snippet of Cleaned Dataset

	A	B	C	D	E	F	G	H	I
1	name_age	match_name	full_name	marathon	year	date	time	time_mini	vaporfly
3	Aaron Braun (M30)	AARON BRAUN	AARON BRAUN	Chicago Marathon	2017	10/8/2017	2:13:41	133.6833	FALSE
4	Braun Aaron (M)	AARON BRAUN	AARON BRAUN	Chicago Marathon	2018	10/7/2018	2:13:16	133.2667	FALSE
5	Aaron Braun (M27)	AARON BRAUN	AARON BRAUN	Houston Marathon	2015	1/18/2015	2:12:54	132.9	FALSE
6	Abayneh Ayele (M28)	ABAYNEH AYEL	ABAYNEH AYELE	Chicago Marathon	2016	10/9/2016	2:13:52	133.8667	FALSE
7	Abayneh Ayele (M)	ABAYNEH AYEL	ABAYNEH AYELE	Houston Marathon	2019	1/20/2019	2:11:30	131.5	FALSE
8	Abayneh Ayele Woldegi	ABAYNEH AYEL	ABAYNEH AYELE W	Houston Marathon	2017	1/15/2017	2:12:44	132.7333	FALSE
9	Abdi Abdirahman (M40)	ABDI ABDIRAH	ABDI ABDIRAHMAN	Boston Marathon	2017	4/17/2017	2:12:45	132.75	TRUE
10	Abdi Abdirahman (M41)	ABDI ABDIRAH	ABDI ABDIRAHMAN	Boston Marathon	2018	4/16/2018	2:28:18	148.3	TRUE
11	Abdi Abdirahman (M42)	ABDI ABDIRAH	ABDI ABDIRAHMAN	Boston Marathon	2019	4/15/2019	2:18:56	138.9333	TRUE
12	ABDI ABDIRAHMAN (M3	ABDI ABDIRAH	ABDI ABDIRAHMAN	New York City Marath	2016	11/6/2016	2:11:23	131.3833	FALSE
13	Abdi Abdirahman (M40)	ABDI ABDIRAH	ABDI ABDIRAHMAN	New York City Marath	2017	11/5/2017	2:12:48	132.8	TRUE
14	Abdi Abdirahman (M42)	ABDI ABDIRAH	ABDI ABDIRAHMAN	New York City Marath	2019	11/3/2019	2:11:34	131.5667	TRUE
15	Abdi Nageeye (M)	ABDI NAGEEYE	ABDI NAGEEYE	Boston Marathon	2016	4/18/2016	2:18:05	138.0833	FALSE
16	Abdi Nageeye (M29)	ABDI NAGEEYE	ABDI NAGEEYE	Boston Marathon	2018	4/16/2018	2:23:16	143.2667	TRUE
17	Abel Kirui (M34)	ABEL KIRUI	ABEL KIRUI	Chicago Marathon	2016	10/9/2016	2:11:23	131.3833	FALSE
18	Abel Kirui (M35)	ABEL KIRUI	ABEL KIRUI	Chicago Marathon	2017	10/8/2017	2:09:48	129.8	TRUE
19	Kirui Abel (M)	ABEL KIRUI	ABEL KIRUI	Chicago Marathon	2018	10/7/2018	2:07:52	127.8667	TRUE

Data Description

This cleaned and merged dataset, split by gender, is comprised of 296 men who ran a total of 862 races, and 270 women who ran 778 races. The average time for men's results in the dataset is 2:18:39, and the average women's is 2:39:19. As Figure 17 suggests, there does not appear to be any overall trends in athlete performance for men or women over time; that is to say that, for this dataset of athletes, men's and women's marathon times have not appeared to be generally getting faster or slower over time.

Figure 17: Average Times by Course and Gender

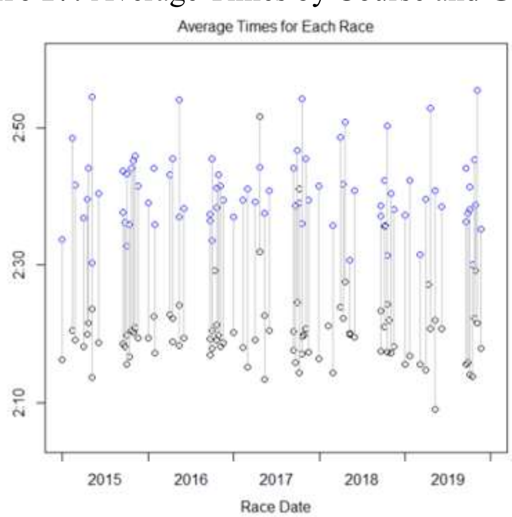
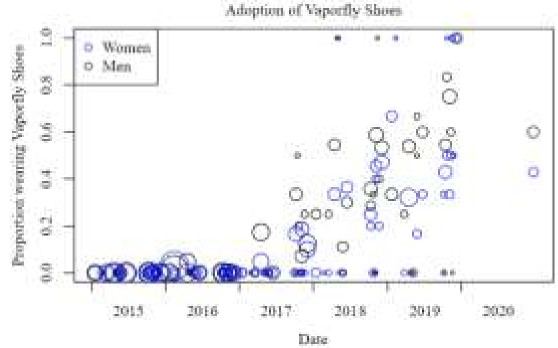


Figure 17 displays the average women's time in blue and average men's time in black for each race, connected by a line to connect the same races for men and women. Out of the 862 male results, 115 of those results were run in Vaporflys, making up 13.34% of the results. More interestingly, the percentage of the results in which Vaporflys were worn after the shoes became available in July 2017 was 36.39% for men, evidence of a large shift towards the shoes in the years after the release of the shoes. Similarly, of the 778 women's results, 85 were in Vaporflys, representing 10.93% of all results. Vaporflys were worn in the women's results 26.81% of the time after they became available to runners in 2017, which, while still representing a large shift towards the shoes once they became available, was not as dramatic of a shift as for the men. This smaller shift on the women's side could potentially be due to fewer of the elite female marathoners in the United States being sponsored by Nike, as other brands such as New Balance and Hoka One One have experienced greater success in attracting female athletes for sponsorships. That remains only a suggestion, though, as only a fraction of the runners in the dataset are professionally sponsored by shoe companies. Further, while the Vaporflys were released in July 2017, adoption was not immediate, as one would expect with any form of technology. Figure 18, a replication of the graph generated by Guinness et. al. in Figure 7, demonstrates the adoption trends of the shoes for men and women:

Figure 18: Replication of Guinness' Adoption of Vaporflys over Time



This graph shows a trend of increasing adoption, as the proportion of athletes in any given race wearing the shoe appears to trend generally positively over time. This trend alone suggests that, even if the effect is not as strong as Nike advertises, athletes believe that the shoes do provide some advantage over other shoes; this is a notable shift among the athletes that depend on running shoe technology the most, as paychecks are on the line at many races for these elite runners.

Methodology & Approach

With data collected and cleaned, analysis could begin. Seeing as the goal of this analysis is to replicate the analysis that Joe Guinness used in his repository, the model specifications are limited to the models described in the paper by Guinness et. al. In short, Guinness employs a restricted maximum likelihood linear mixed model. A linear mixed model is used, as opposed to a simple linear regression model, in order to account for mixed effects, random or fixed. In this scenario, the primary areas of interest are the fixed effects presented within, first, distinct runners, and second, distinct courses in distinct years. The two specifications, one in normal form and the other in log form, are given below:

Figure 19: Model Specifications

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x1 + (1 | f1) + (1 | f2)
Formula: log(y) ~ x1 + (1 | f1) + (1 | f2)
```

Here, *y* represents time in minutes for a given performance, *x1* represents the Vaporfly dummy variable (True or False, 1 or 0), *f1* represents runner-specific fixed effects, and *f2* represents course (see Appendix) and year (2015-2019) specific fixed effects. While relatively simple, this model is powerful; by controlling for runner-specific fixed effects, the model attempts to control for a runner's baseline performance level, while the year and course-specific fixed effects attempt to control for a course's difficulty as well as the conditions of the race in a given year. These fixed effects allow the model to strip away other key factors that influence runner performance, and subsequently hone in on the effect produced by the Vaporflys.

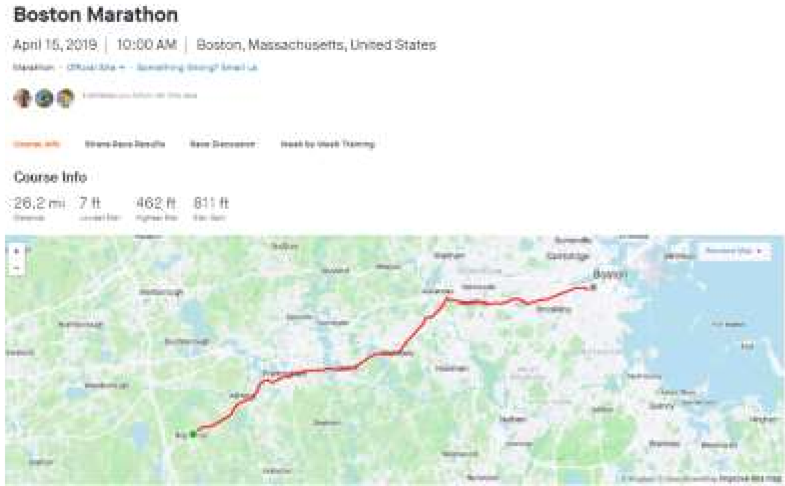
Strava Analysis

Data Needs

For the next section of analysis, the focus shifts from professional and elite athletes to amateur runners. Taking inspiration from the *New York Times'* analysis of the Vaporfly effect, Strava is used for large quantities of publicly available marathon race data. Instead of limiting the races to competitive North American marathons, the six World Marathon Majors are used: Berlin, Boston, Chicago, London, New York and Tokyo. Using results from 2014-2019, these races attract a wide variety of runners from across the world. In order to build a variety of models for analysis, most of the publicly available Strava activity data is collected from these races. A Strava activity is simply an upload record from an individual's run. It includes, at the minimum, their time, distance, pace, course and elevation change. Runners can also choose to report their shoes and gear used during the activity on their post.

First, data from Strava's race pages is collected. Race pages, such as the one displayed in Figures 20 and 21, include tabulated results with the race course and date, athlete name and ID, age, gender, finish time and pace, and the athlete's Strava activity title and ID. All of these fields provide valuable information for analysis and can be easily collected using program scripts to manipulate the underlying HTML.

Figure 20: 2019 Boston Marathon Race Page Header



Source: strava.com/running_races

Figure 21: Top 10 Results on Strava, 2019 Boston Marathon

Source: strava.com/running_races

Then, using the activity ID's from a race page, it is possible to collect further information about athletes' individual performances. This information includes heart rate and Strava's suffer score data, a measure of exertion based on heart rate data, any information the athlete writes in the description box of the activity, and, most notably, what shoes the runner was wearing during the race.

Figure 22: Sample Result from Chicago 2019, a Personal Best

Figure 23: Sample Result from Chicago 2019 Including Shoe Data

Source: strava.com/activities

Figures 22 and 23 demonstrate the additional information that can be collected from an individual athlete's activity. Useful pieces of data, such as whether or not the performance was a personal best for the athlete, can often be found in these activity pages. By combining these two rich data sources for each race, a dataset can be prepared for a variety of analyses.

Data Collection

Recognizing that the data exists and is publicly available is simple; collecting large amounts of that data in a way that will be conducive to econometric analysis is entirely more challenging. So, the next step in the process is building the dataset. With the goal of building a large dataset with tens of thousands of results, manual collection of any of this data is immediately ruled out. The entire process would need to be automated to be performed by a program script. Thus, Python programs are developed by taking inspiration from the Python scripts used in the Guinness scraping process and adjusted to handle the new scraping environments.

The first step in this scraping process is collecting the desired race results by visiting Strava race pages. The first challenge here is determining which races to scrape, and how to access those races' race page on Strava's website. At first glance, it appeared that the URLs for race pages took the form of ‘strava.com/running_races/[RACE_ID#],’ where the RACE_ID# is some number generated by Strava that is meaningless other than being a unique identifier for a given race page. After further inspection, certain races have alternative URLs that take the following form: ‘strava.com/running_races/[YEAR]-[RACENAME]’ (ex: https://www.strava.com/running_races/2014-Boston-Marathon). This format is more conducive to automating the scraping process, as this format would allow for compiling of a list of marathon race names and years, and simply iterate through those two lists to collect all of the race results from those courses and those years. For simplicity and to ensure comparable athlete levels, the list of races is restricted to the six World Marathon Majors: Boston, Chicago, New York, Tokyo, London and Berlin. Strava first started creating race pages for the 2014 Boston Marathon, so scraping starts with the year 2014 and goes through 2019, the latest year in which any of these marathons were run. Because some race pages are broken or missing, the final dataset includes a total of 30 races from those six marathons over the six-year time span, 2014-2019.

From a technical perspective, a variety of Python packages are used to visit the race pages, scrape them for the appropriate data, and then store the data in a usable format. First, upon iterating over the list of races and years, the Python requests package was used to automate the visiting of a page and retrieval of the page’s HTML. Once on the results page, BeautifulSoup, another Python package, was used to inspect the HTML of the page and collect the desired information based on HTML tags. On each results page, such as the one displayed in Figure 21, all information available in the results table is collected: athlete name, athlete ID, activity name, activity ID, time, place, gender, and age range. Because only 20 results are visible at a time on a race page, the program iterates through the results table until all results are collected or 2,000 results are collected, whichever comes first. The maximum number of results collected per page is capped a 2,000 for primarily practical and convenient reasons: if all results were collected, the dataset would balloon in size so much that collecting specific athlete data would be more arduous than it already is, as will be explained later. Once the BeautifulSoup package collects all of the relevant race page data by inspecting HTML tags, the data is stored in a Numpy array for saving in CSV format at the end of the scrape. After this first scrape of the 30 different race instances was complete, the dataset reaches a size of over 52,000 results, or about 1,700 per race.

Before moving to the next key stage of data collection, the dataset is reviewed. While a large dataset can provide for more compelling results after analysis, the vastness of the dataset poses practical challenges to the next stage in the data collection process. The first segment of collection featured visiting only 30 pages and collecting information from those pages, but the next segment entails visiting all 52,000+ activity pages for each athlete's individual activity for a given race. Even though it is automated, scraping still takes time, so the dataset needs to be pared down. Thus, results are only kept for athletes that run at least two marathons in the present dataset; this allows for comparisons and analysis between races for the same athlete. After eliminating athletes with only one result in the dataset, there remain roughly 25,000 activities to scrape from just over 9,000 athletes. Table 1, below, displays the number of results coming from each marathon each year and each race's average time from their respective years.

Table 1: Race Year Result Count and Average Time

race and year	Freq.	mean(timeMin)			
BERLIN			LONDON		
2014	372	208.7949	2014		
2015	700	196.6806	2015	735	183.4654
2016			2016	1,040	177.7788
2017	930	180.8044	2017		
2018	1,034	179.8959	2018	1,146	181.9761
2019	889	170.9584	2019	1,064	169.5546
BOSTON			NYC		
2014	608	197.7186	2014	547	221.8368
2015	1,035	190.896	2015	796	207.4763
2016	1,257	188.9536	2016		
2017	1,329	183.3302	2017	1,096	187.9495
2018	1,285	178.8578	2018	1,134	183.0305
2019	1,187	172.2929	2019	1,001	178.2704
CHICAGO			TOKYO		
2014	493	215.6692	2014		
2015	698	215.5815	2015	119	226.5172
2016	932	196.0008	2016	24	233.9812
2017	1,090	189.6206	2017	364	216.3895
2018	1,163	181.595	2018	400	209.4172
2019	983	174.5422	2019		

The next section of scraping is more difficult than the first. At first glance, because the activities are public records on Strava, it seems that the process is similar to the previous. However, after further investigation, it becomes clear that data on the shoes that the athlete wore during an activity is not available by simply visiting activity pages. While these pages are publicly accessible, the data of interest is hidden behind a log-in wall. For that reason, a logged-in browser would need to visit and scrape the pages. There are a variety of Python packages available that are capable of doing this; after playing around with packages like mechanize and robobrowser, MechanicalSoup proved to be the package of choice. This package combines the capability of mechanize, which allows for advanced page interaction, such as completing the authentication required in the log-in sequence, while also handling the HTML similarly to BeautifulSoup.

While a good solution to the problem of hidden data, MechanicalBrowser is not as fast or efficient as the simpler requests package. The process of logging in and then navigating to pages from this page is very time consuming, and, once compounded with the need to visit 25,000 pages instead of just 30, completing this scrape becomes a multi-day process. In addition to the slower visiting and scraping process, Strava's site request limits prevent smooth scraping. A logged-in user is limited in the number of times they can make a request to the Strava servers in fixed time-frames; for example, Strava lists those limits to be 100 requests per hour and 1000 requests per day on their developers' page. In order to expedite the process, four different Strava user accounts are used. Different methods were tested, such as scraping with all four accounts simultaneously on different computing machines, cycling between accounts and gaming the request limits. Finally, after about 5 days and well over 40 hours of scraping, the activity-specific data had been collected for the 25,451 race results in the dataset. Cross-sections of that scraped data are as follows in Figure 24. This data is not yet merged with the master result data, limiting the fields that are shown:

- *activity_id* is the unique numerical identifier for a given Strava activity
- *shoes* details the shoes an athlete wore, if reported; otherwise, '\xe2\x80\x94' fills if value is blank
- *device* reports the gear used to record an activity, such as a specific GPS watch; '\xe2\x80\x94' fills if value is blank
- *suffer* records Strava's heart rate-based measure of exertion; '\xe2\x80\x94' fills if value is blank

Figure 24: Sample of Scraped Strava Data

1	activity_id	shoes	device	suffer
2	1052659650	\xe2\x80\x94	Garmin f\xc4\x93nix	\xe2\x80\
3	1666007311	\xe2\x80\x94	Garmin Forerunner 9	359
4	739742222	Brooks Pureflow\n(322.6 mi)	Garmin Forerunner 2	733
5	2326841948	adidas Boston 6\n(3,268.3 mi)	Suunto Ambit3 Peak	421
6	549394289	\xe2\x80\x94	Garmin Forerunner 2	\xe2\x80\
7	289236555	New Balance 890v4 Boston 2nd	Garmin Forerunner 6	563
8	2292546299	\xe2\x80\x94	Garmin Forerunner 2	452
9	2838932874	Nike Vaporfly Next %\n(52.6 mi	Garmin Forerunner 6	\xe2\x80\
10	1946595254	Nike Zoom Fly Black\n(4,058.5 r	Garmin v\xc3\xadvoa	\xe2\x80\

With all of the appropriate data scraped and output into CSV files, the datasets are ready for cleaning and merging. Further Python scripts are used to clean up unfavorable formatting and unusual values in each dataset. An example of unusual values is any field that was filled with '\xe2\x80\x94,' a placeholder for null values. The programs loaded the CSV data into Pandas data frames and then iterated over certain columns to clean data by stripping strings of UTF-8 encodings and converting strings to integers. Then, the program generates new variables that are important for later analysis by manipulating the given cleaned data. Most notably, the program creates the *vaporfly* variable by iterating through the *shoes* variable from the activity dataset and searching that string for sub-strings such as "4%," "Vaporfly," "NEXT%," or "Alphafly," all used in the naming of the shoe, and giving value 1 where present or 0 where not. Other variable generation based on the existing dataset included values for *race*, *year,* and *raceYear*, *selfReported_PR* and *strava_PR*, and *z1suffer*. The *race, year* and *raceYear* variables capture the course the result was run on, the year in which the result was run, and the interaction of those two terms *race * year*, respectively. The *selfReported_PR* variable is generated based on any indication by an athlete in their activity title that the race was a personal record, while the *strava_PR* variable takes the value 1 for the races that are an athlete's best performance in the dataset and 0 otherwise. The *z1suffer* variable is a standardization of the suffer score variable around its mean, after unusually high or low values are dropped to avoid any bias due to outliers. This standardization is performed by subtracting the mean of *suffer* from each instance of *suffer* and then dividing that value by *suffer*'s standard deviation. Ultimately, the dataset has missing values in places where athletes did not report certain pieces of information, such as the shoes they wore or their average heart rate. Those missing values are left blank in the dataset, and then handled at time of analysis. Finally, after each of the two datasets are cleaned, they are merged by matching activity data to results data based on the *activityID* listed in both datasets. A sample of the final dataset follows in Figure 25.

Figure 25: Sample of Final Dataset, with a Variety of Shoes & some Missing Values

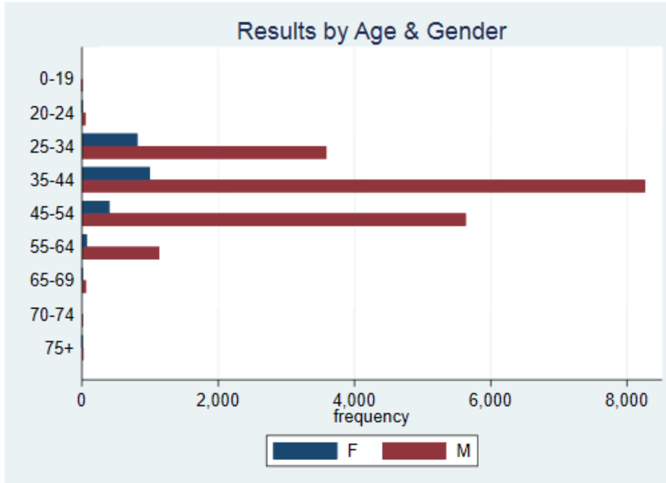
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	athlete_id	Name	Gender	Age	ActivityTitle	Finish	Pace	data-activity_ic	athleteflag	year	race	timeSec	vaporfly	shoes	suffer	reported_	dataset_b
443	7374175	Adrian B.	M	55-64	b'04/20/2015 Hopk	3:15:29	4:38/km	290047664	2	2015	BOSTON	11729	0			0	0
444	7374175	Adrian B.	M	55-64	b'London marathor	2:58:40	4:14/km	2326244216	2	2019	LONDON	10720	1	Nike Vaporfly 4%\n(		0	1
445	7374606	Adrian Catherwic	M		b'London Marathor	2:59:31	4:15/km	555430410	2	2016	LONDON	10771	0	Mizuno Blue Wave C		0	1
446	7374606	Adrian Catherwic	M		b'Morning Run'	3:21:57	4:47/km	1553519435	2	2018	LONDON	12117	0	Mizuno Again Wave		0	0
447	19154383	Adrian Cooper	M	45-54	b'London Marathor	3:21:54	4:47/km	1524620424	2	2018	LONDON	12114	0			0	0
448	19154383	Adrian Cooper	M	45-54	b'London Marathor	2:58:34	4:14/km	2325335729	2	2019	LONDON	10714	0			0	1
449	14914537	Adrian Cruz	M	35-44	b'TCS NYC Maratho	3:01:39	4:18/km	1946205235	3	2018	NYC	10899	1	Nike Vaporfly 4% Flyl		0	1
450	14914537	Adrian Cruz	M	35-44	b'TCS NYC Maratho	3:11:27	4:32/km	2838849032	3	2019	NYC	11487	1	Nike Vapo	302	0	0
451	14914537	Adrian Cruz	M	35-44	b'Chicago Maratho	3:02:34	4:20/km	1890387257	3	2018	CHICAGO	10954	0	Nike Pegasus 35 Turl		0	0
452	5213276	Adrian Eckhardt	M		b'41. Berlin Marath	3:11:10	4:32/km	200413442	2	2014	BERLIN	11470	0			0	1
453	5213276	Adrian Eckhardt	M		b'44. BERLIN-MARA	3:21:48	4:47/km	1199916864	2	2017	BERLIN	12108	0			0	0
454	15582094	Adrian H.	M	25-34	b'Boston Marathon	2:41:47	3:50/km	2292146140	2	2019	BOSTON	9707	0	Nike Zoom	332	0	0
455	15582094	Adrian H.	M	25-34	b'Chicago Maratho	2:25:30	3:27/km	2786062489	2	2019	CHICAGO	8730	1	Nike Vapo	262	0	1
456	1853437	Adrian Hall - Jou	M	45-54	b'04/20/2015 Hopk	3:49:33	5:26/km	290805340	2	2015	BOSTON	13773	0	HOKA ONE ONE Bon		0	1
457	1853437	Adrian Hall - Jou	M	45-54	b'11/02/2014 State	3:54:11	5:33/km	222050449	2	2014	NYC	14051	0	HOKA ONE ONE Bon		0	0

This cross-section of the final merged dataset shows a number of performances from athletes named Adrian. One can see how certain fields, such as *shoes* and *suffer*, are recorded in some activities and left null in others. This cross-section also features multiple performances in which athletes run their fastest time in the dataset while wearing the Vaporflys.

Data Description

The dataset is made up of 25,451 unique results run by 9,633 unique athletes. Males make up the majority of the race results, accounting for 22,151 of the results, or just over 90% of the dataset. The Boston Marathon is the most popular race in the dataset, accounting for 26.33% of results; Berlin, Chicago, New York City and Chicago all represent similar shares in the 15-20% range, while Tokyo accounts for only 3.56% of the results in the dataset, reflecting not the actual size of the race, but the adoption of Strava among finishers at Tokyo. Additionally, the majority of results come from athletes that land in the 35-44 age group, making up 9,288 of the dataset, as shown in Figure 26.

Figure 26: Results by Age Group & Gender



The average result for men in the dataset is 3:05:31, while for women the mean is 3:20:57. More interesting, though, is how those results have progressed over time in the dataset.

Figure 27: Average Time by Course and Year

Figure 27, above, depicts the trends in result times by race and year. These line graphs clearly show that times have gotten faster in the dataset over the 2014-2019 range. This trend breaks with the trend seen in the Guinness et. al. replication, where there were no noticeable trends in average time. Table 2, below, presents a more detailed view of how these results change over time in the dataset, along with Vaporfly adoption and change in suffer score over time.

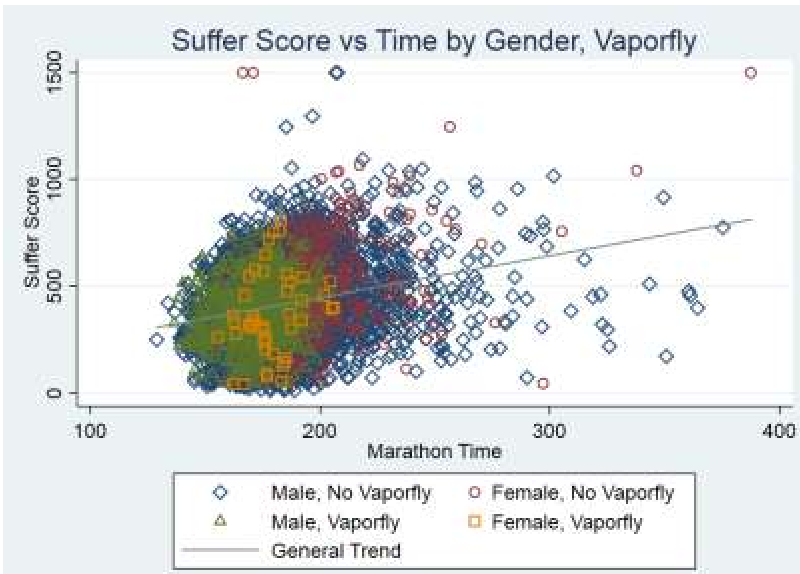
Table 2: Average Time, Vaporfly Adoption and Suffer Score by Gender and Year

Gender and year	Freq.	mean(timeMin)	mean(vaporfly)	mean(suffer)
F				
2014	336	227.0764	0	547.72729
2015	555	215.266	0	522.64581
2016	386	199.9747	0	521.88892
2017	536	197.074	.01679105	477.87878
2018	542	190.5874	.0498155	433.85715
2019	391	178.8223	.16112532	429.37167
M				
2014	1,657	207.2259	0	476.32169
2015	3,451	196.2701	0	455.63455
2016	2,812	186.0221	0	437.48682
2017	4,143	186.5356	.01593048	418.02487
2018	5,467	182.0328	.06932504	398.74112
2019	4,621	172.58	.19649427	384.94556

As the above table demonstrates, for both men and women, the average result is faster year over year from 2014 to 2019. The number of results in a given year, does not follow the same trend; while 2014 has the lowest number of results for both genders, the number of results varies widely by year after that. This variation in number of results could be attributed to certain popular marathons, such as New York City and London, not having 2016 or 2017 results pages on Strava, respectively. However, Strava, a relatively new platform, has seen massive growth over that same time frame, adding millions of users in that date range. Thus, one might hypothesize that the downward trend in average result has to do with the collection method; because only the top 2,000 results are collected from each race, an increase in results reported on Strava per race would likely cause the top 2,000 results to get faster each year. Similarly, athletes' suffer scores, a measure of heart rate, has been trending downward through the dataset, likely a trend induced similarly by the collection method. Thus, this collection method could introduce a certain degree of selection bias to the analysis. When analyzing results, these general upload trends on Strava's platform needs to be handled so that selection bias does not impact estimates of the Vaporfly effect.

Despite the concerns surrounding selection bias, Table 2 presents intriguing trends with respect to marathon times and suffer scores. In Figure 28, potential relationship between time, suffer score and the Vaporflys is explored by gender.

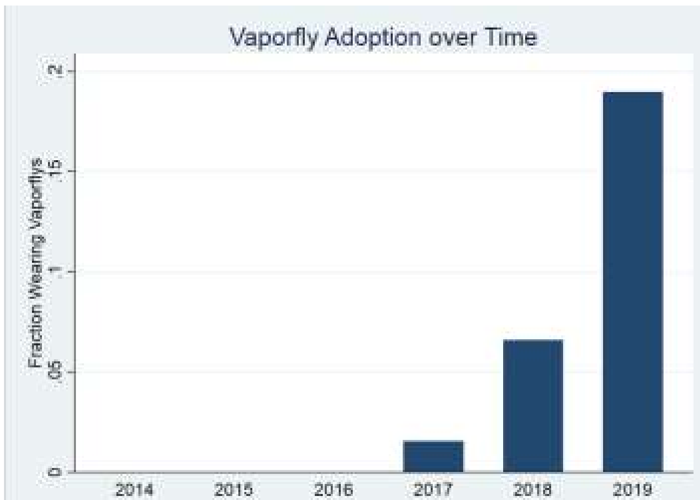
Figure 28: Suffer Score vs Time by Gender & Vaporfly



In this scatterplot, a few interesting trends appear. Not surprisingly, athletes with lower suffer scores are generally running faster than runners with higher suffer scores; suffer score increases with time and vice versa. More interestingly, though, is the fact that nearly all results, male or female, by a runner in Vaporflys are aggregated around both lower suffer scores and faster times. This trend has complicated implications. If the suffer score is used as a measure of fitness, does this trend then imply that fitter athletes are also wearing Vaporflys, or does it imply that the Vaporflys result in lower exertion, as represented by the suffer score? This figure is a first look at potential complications that could be associated with findings later on.

In addition to demonstrating trends in results, Table 2 gives insight to adoption of the Vaporflys over time. Figure 29, below, demonstrates how adoption of the shoe among runners in the dataset has been increasing quickly over time. By 2019, nearly 20% of runners in the dataset are using the Vaporflys in their race.

Figure 29: Vaporfly Adoption over Time



As expected, the fraction of runners wearing Vaporflys is 0 before their release in 2017. Subsequently, the fraction of adopters of the shoes increases each year, with both men and women starting with about 1.5% adoption in 2017 results, increasing all the way to over 16% of female results in 2019 and nearly 20% of male results in 2019. This trend mirrors the trend among elites in the Guinness study, showing that amateurs are also shifting towards the shoe in greater numbers each year.

Men wearing the Vaporflys ran 15 minutes faster on average than men who did not wear Vaporflys, and women wearing the shoes averaged 22 minutes faster than their counterparts who did not wear the shoes. However, those averages do not account for time trends of results in the dataset; times are trending faster in the dataset even before the release of the Vaporflys. Thus, Figure 30 shines light on how this trend and the Vaporfly effect can potentially be unified:

Figure 30: Results Trends by Vaporfly & Gender

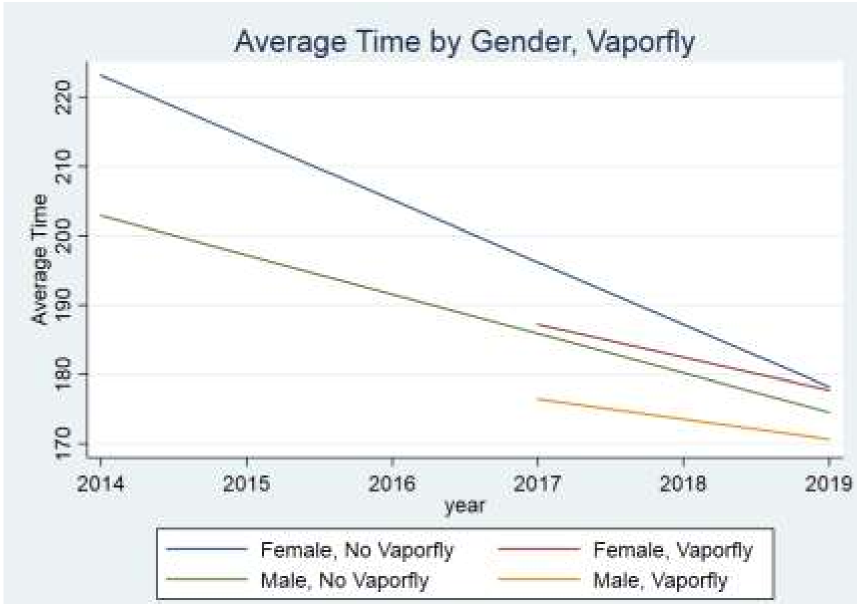


Figure 30 demonstrates how times in our dataset improve for both Vaporfly-wearers and those who are not wearing the shoes. Interestingly, Vaporfly-wearers' results converge with non-wearers' results for women, calling into question the Vaporfly effect. Meanwhile, for men, the results trends for wearers and non-wearers of the Vaporflys are nearly parallel. The advantage for the wearers of the shoes remains nearly constant over time, building a graphical case for a Vaporfly effect. While these simple summary statistics are certainly suggestive of the Vaporflys providing a measurable advantage in race performance, they should be interpreted as nothing more than just that, suggestions.

Methodology & Approach

With the bulk of the data collected, processed, and merged into clean datasets and exploratory analysis complete, estimating the Vaporfly effect becomes a question of the strength of econometric analysis. While in the Guinness replication the goal was to directly recreate the analysis presented in their paper, the goal of the Strava analysis will be to iterate through a variety of model specifications, building each model on the last, justifying each added feature.

At a high level, our models for analyzing the Strava data will have two primary forms that are then split by gender, creating a total of four total models that will be run for any subsequent specification:

Data Split #1: timeMin vs logTime

Those two forms will be an untransformed regression analysis, in which the dependent variable, *timeMin*, or marathon time in minutes, is left untransformed, and a log-transformed regression analysis, in which *timeMin* is converted to *logTime* by taking the natural log of all of the results in the dataset. The *logTime* dependent variable is of primary interest in this case because the regression analysis will subsequently yield coefficients that estimate the percentage change in time attributed to each explanatory variable in the model. Estimating these outcomes in percentage format then allows for a meaningful comparison with Nike's claim that the shoes will make an athlete 4% faster.

Data Split #2: Gender

The models are then split by gender, male or female. This is done for a variety of reasons but, most prominently, because other studies of the shoes have handled gender similarly; other studies have found measurable differences in the Vaporfly effect by gender, and being able to compare the results found here to other studies is valuable.

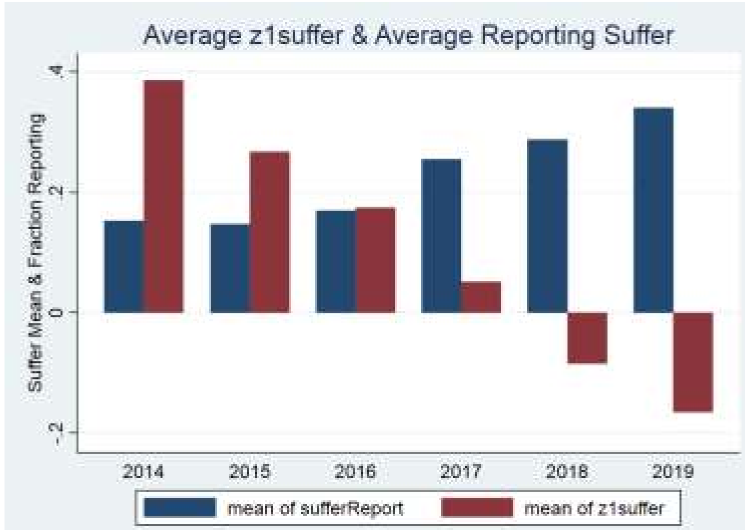
The basis for this regression analysis is relatively simple: a simple linear regression with *logTime* or *timeMin* as the dependent variable and *vaporfly* as the primary explanatory variable of interest. As a result, the *vaporfly* variable's coefficient estimated in the regression analysis will ultimately be the estimated effect of wearing the shoes, or the Vaporfly effect, controlling for everything else in the model.

The first iteration on the base model is a relatively simple multi-linear regression, using z-score standardized suffer scores as a second explanatory variable, as follows:

$$logTime = b0 + b1*vaporfly + b2*z1suffer + e$$

By adding the standardized suffer score variable as an explanatory variable, changes in marathon times are explained by a proxy for that athlete's fitness at the time of the race. The heart rate data variable, *suffer*, provides an estimate of fitness level because we expect that a runner that completes a run at a lower heart rate has a higher fitness level than a runner that completes the run with a higher heart rate. Because not all athletes on Strava report their heart rate data in their activity, including this variable restricts the dataset. This variable is standardized in this model in an attempt to assign more explanatory power to the variable. Now centered around 0 with standard deviation of 1, any *z1suffer* that is less than 0 represents activities with lower average heart rates and thus lower exertion, while values greater than 0 implies greater exertion. While reducing the number of observations when included in the model, this method adds a second meaningful explanatory variable, a benefit that might outweigh the cost of reducing the dataset in size. Figure 31 presents trends in the reporting of suffer score as well as the measure of the standardized suffer variable, *z1suffer*, over time. The suffer score variable has followed interesting trends over time in the dataset. Reporting over the variable has increased over time, from close to 15% in 2014 to over 30% in 2019, likely reflecting higher rates of heart rate monitor usage by athletes over time.

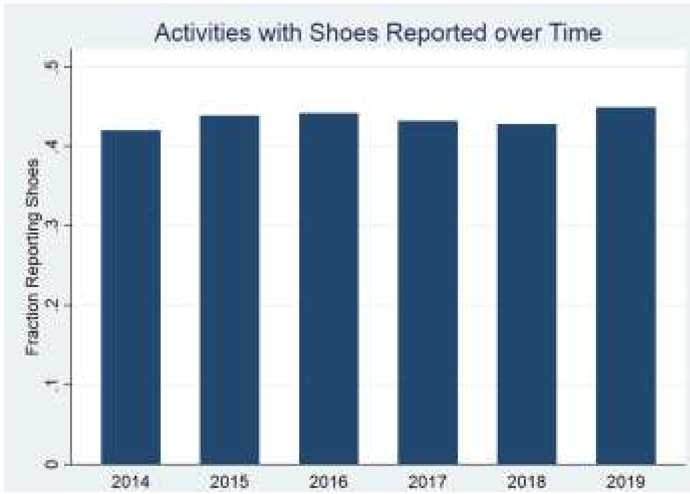
Figure 31: Average *z1suffer* & Fraction Reporting *suffer*



This increase in reporting is accompanied by a decrease in suffer score in activities. This trend is cause for concern, as it may suggest that suffer scores are not decreasing because of improved fitness, but instead because more athletes have access to the heart rate monitor technology. For our *z1suffer* variable to have good explanatory power, it should not be correlated with the proportion of people reporting that measure, as that introduces a systematic bias to the variable.

The restrictions of our dataset as a result of missing *z1suffer* values raise an important consideration in estimating the Vaporfly effect using the *vaporfly* variable as well. Like with heart rate data, not all athletes report the shoes they wore in their activity; for these activities, *vaporfly* is assigned a value of 0. However, it is possible that an athlete might wear the shoes but still not report it on Strava. Figure 32 displays the fraction of runners reporting the shoes they wore over time in the dataset. Year to year, activities reporting shoes varies between 41% and 45% with no visible trend in reporting. This recognition gives rise to another variation that will be useful to run on each subsequent model specification. This potentially errant assumption can be avoided by restricting analysis to activities that are not missing reported shoes.

Figure 32: Fraction of Activities Reporting Shoes



While this restriction to the roughly 42% of activities in which shoes are reported reduces dataset to about 9,000 activities, potentially reducing the empirical explanatory power, it removes another potential source of bias in the estimates.

With those relatively simple variations on linear regressions formulated, now models can be developed that handle greater sources of bias through the use of fixed effects. Fixed effects are used to capture the average quality of a variable not explained by the rest of the model, so, by adding them to the trends in the linear regression models already built, the effects of those variables can be stripped from the estimate of the Vaporfly effect. Fixed effects will be used for three variables: athlete-specific effect, course-specific effect, and race-year-specific effect.

- First, the athlete-specific effect can be used to set a baseline talent level for each athlete in the dataset, represented by the *athlete_id* variable. Estimating this fixed effect helps to differentiate outcomes that differ as a result of one athlete being naturally faster or more talented than another.

- Next, the course-specific fixed effect accounts for the average difficulty of a course. It is not necessarily fair to compare the New York City or Boston marathons, considered relatively hilly and challenging courses, to the flat and fast courses in Chicago or Berlin. This fixed effect accounts for those differences in difficult level and removes it from the estimated Vaporfly effect.
- Finally, the race-year-specific fixed effect is used to control for the conditions on race day at a particular race. While this effect may seem trivial to some, the effect of weather on a performance should not be underestimated. Take the Boston Marathons run in 2018 and 2019, for example. In 2018, freezing rain and gusting wind produced two of the slowest winning times for the race in decades. Meanwhile, in 2019, results were as fast as ever on a perfectly sunny, mild 50-degree day.

A model controlling for each of these fixed effects might be presented as follows:

$$logTime = b0 + b1*vaporfly + b2*z1suffer + (1|f1) + 1(|f2) + (1|f3) + e$$

where *f1*, *f2*, and *f3* are fixed effects captured by the *athlete_id*, *race,* and *raceYear* variables, respectively. With this more complex fixed effect model in place, all previously mentioned variations can also be run: including or excluding *z1suffer*, restricting to activities that reported shoes, using untransformed or log-transformed dependent variable, and separating analysis by gender.

Results

Guinness Replication

The primary goal of this analysis is any improvement, or effectively replication, upon Guinness' results by way of adding more data to the dataset. The additional data in this dataset is the 2020 Marathon Project, a race run at the end of 2020 in order to give professional marathoners a chance to compete during the COVID-19 pandemic. The results are as follows, separated by gender:

Results 1: Guinness Improvement, Men

Guinness Replication Men

	(1) time_minutes		(2) logTime	
main				
vap	-3.060***	(-5.98)	-0.0218***	(-6.13)
_cons	138.8***	(473.35)	4.932***	(2356.40)
lns1_1_1				
_cons	1.375***	(22.91)	-3.550***	(-60.27)
lns2_1_1				
_cons	1.414	(0.33)	-3.559	(-0.73)
lnsig_e				
_cons	0.545	(0.02)	-4.412	(-0.17)
N	862		862	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Results 2: Guinness Improvement, Women

Guinness Replication Women

	(1) time_minutes		(2) logTime	
main				
vap	-2.115**	(-3.03)	-0.0133**	(-3.08)
_cons	158.9***	(363.96)	5.067***	(1833.89)
lns1_1_1				
_cons	1.812***	(31.57)	-3.242***	(-57.14)
lns2_1_1				
_cons	1.538	(0.25)	-3.551	(-0.53)
lnsig_e				
_cons	0.762	(0.03)	-4.317	(-0.14)
N	778		778	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

As the estimated coefficient on *x1* indicates, this model produces a slightly stronger estimated Vaporfly effect for men than reported in the original Guinness et. al. study, while estimating a roughly equivalent effect for women. These differences are interesting nonetheless: the estimated effect for men is now a 2.18% performance boost, compared to 2.09% in the original analysis, and for women a 1.33% performance boost, compared to 1.35% in the original analysis. Despite the addition of new data from a highly competitive race, the estimates do not fluctuate much from the original analysis, further confirmation of Guinness et. al.'s estimate under the specified model.

For men, this advantage represents an advantage of over three minutes in a marathon race, while for women the advantage is over two minutes, both sizable margins at the professional level. Most professional marathoners would be pleased with a 30 second improvement on their personal best; these shoes promise the possibility of an improvement of over 4-6 times as strong. With a large number of observations for both the men and the women, and high t-values for each regression, the empirical value of each of these results is promising.

While these results seem more reasonable at face value than Nike's claim of a 4% improvement, they still have holes with respect to omitted variable bias. While these models control well for a variety of fixed effects, they do not do as good of a job at accounting for any effect that might be attributed to improvement in fitness over time. While athlete-specific fixed effects control for an athlete's general talent level, there are still fluctuations in how fit an athlete might be in any given performance. On one hand, a certain race result might be slower than an athlete would expect to run if he or she had been plagued by injuries during the training cycle. On the other hand, a result could also represent a breakthrough performance after years of quality training. This model makes no attempt to control for either of those factors that could very easily sway the estimate for the Vaporfly effect. One might also hypothesize that the models also do not account for the effect that might be attributed to wearing a fresh pair of shoes, as Vaporflys are generally only worn for 3-4 races, as opposed to a pair of shoes that might have hundreds of miles of training on them, but we can assume that this effect is null in this analysis because the nature of professional contracts is such that athletes are able to wear a new pair of shoes in each passing race if they so choose; this assumption does not hold for amateurs, however.

Nonetheless, the results from this replication are still compelling. While the accuracy of the estimated effect is still up for debate, the existence of such an effect is reaffirmed. At such large scale, it seems hard to believe that the estimated effect can be attributed entirely to omitted variables. As such, the replication and improvement upon these results achieves the goal of setting a baseline estimate for the Vaporfly effect within the context of professional running.

Strava Analysis

Next, the analysis of Strava data seeks both to quantify the effects for a broader population of runners and to hone in further on the Vaporfly effect by controlling for omitted variable bias not controlled for in the Guinness analysis. As described in the methodology section, an iterative approach is used to achieve these goals, building models from each other. Results for all model variations are presented in results tables 3 and 4:

Results 3: Strava Analysis, Men

Strava Results, Men

	(1) Simple MLR	(2) Simple FE	(3) MLR FE	(4) MLR FE No ~g	(5) FE No Miss~g
vaporfly	-0.0650***	-0.0270***	-0.0211***	-0.0237***	-0.0306***
	(-16.29)	(-12.03)	(-6.28)	(-6.23)	(-11.47)
z1suffer	0.0233***		0.00282*	0.00557***	
	(16.66)		(2.22)	(3.33)	
_cons	9.308***	9.312***	9.300***	9.297***	9.309***
	(6430.76)	(25982.12)	(11985.01)	(7910.70)	(15263.40)
N	5659	22142	4883	2745	9241

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Results 4: Strava Analysis, Women

Strava Results, Women

	(1) Simple MLR	(2) Simple FE	(3) MLR FE	(4) MLR FE No ~g	(5) FE No Miss~g
vaporfly	-0.0619***	-0.0173**	-0.0150	-0.00606	-0.0152*
	(-3.86)	(-2.82)	(-1.51)	(-0.49)	(-2.01)
z1suffer	0.0265***		0.000126	-0.00414	
	(7.58)		(0.04)	(-0.78)	
_cons	9.351***	9.380***	9.343***	9.347***	9.390***
	(2002.26)	(12471.09)	(4439.02)	(2359.63)	(6560.70)
N	553	3297	436	166	949

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Model 1: Simple Multi-Linear Regression

The first regression that attempts to quantify the effect is a relatively simply multilinear regression. The standardized suffer score variable is used as an explanatory variable to stand in for an athlete's fitness during the race. As such, this regression concludes a 6.50% Vaporfly effect for men and 6.19% effect for women—sizable! These results should not be trusted though, as they fail to control for all of those athlete, course and race year specific fixed effects.

Model 2: Simple Fixed Effects

Next, controlling for the aforementioned fixed effects without including suffer score or excluding runners that did not report their shoes on their activities in the model results in regression results with the highest number of observations for both men and women, 22,142 and 3,297, respectively. Upon controlling for the fixed effects, the estimated Vaporfly effect plummets for both men and women. These results now more closely mirror results seen in other studies, with the effect estimated at 2.70% for men and 1.73% for women. However, the goal is not simply to mirror other results, but to improve upon them.

Model 3: Combining simple MLR with FE

In order to further control for the fitness effect that the Guinness study failed to control, these two preliminary models are now combined by running a multi-linear regression with fixed effects. As one might expect, the combination of these two models drives down the estimate even further. Now, for men, the model estimates a 2.11% effect with strong statistical significance, while for women it estimates a 1.50% effect, a value that is called into question by a suspiciously high p-value of 0.132. This high p-value makes this estimate for the women's Vaporfly effect untrustworthy. While this model is as advanced as the specification gets at this point, there still exist questions surrounding a variety of biases and uncertainties.

Model 4: MLR & FE with no Missing Values

In the next iteration on the model, any uncertainty caused by the tagging of the Vaporflys is addressed. Marking activities that did not include any report of shoes as 0, or not using Vaporflys, presents a clear source of bias; while one might anticipate that wearers of the Vaporflys would report their shoes, this assumption is weak and unproven. For that reason, in the next models, only activities that included the designation of the shoes worn in activity are included. By limiting the analysis to this subset of activities, the estimates are again changed, but in different directions for men and women, which also comes across as suspicious. For men, the effect is now estimated at 2.37% on the 2,745 observations still in the model, an increase in estimated Vaporfly effect from the previous iteration despite the decrease in observations, still with strong statistical significance. Meanwhile, the women's estimate drops all the way to a 0.61% effect with verry little statistical significance; this lack of statistical significance, represented by a very high p-value, could be due to the women's sample size dropping down to just 166 activities. Because the sample size is so shrunken, this estimate becomes even less relevant. Focusing on the estimate for men, however, provides an interesting result: when *vaporfly* is assigned with certainty based on non-missing values, the model now estimates a stronger Vaporfly effect.

Model 5: Fixed Effects with no Missing Values

While model 4 seems the most advanced, combining an interesting multi-linear regression with meaningful fixed effects, there remain doubts surrounding the validity of heart rate data as a proxy for fitness. The assumption behind that proxy is that athletes with lower heart rates are fitter than others; this assumption is not perfect. For example, different athletes's heart rates might simply respond differently to exertion levels. Additionally, the trends presented with respect to suffer score reporting and its value trends cause more concern for bias within that variable. This is not to say that this specification should be thrown out all together, but simply to say that it is worthy of some skepticism. For that reason, the following model is run leaving behind the *z1suffer* variable, but still requiring that the activity include the shoes the athlete wore on race day. Here, the model produces the strongest estimated Vaporfly effect yet under the fixed effects model for men, coming in at 3.06% on 9,241 observations. The women's result comes in at an estimated 1.52% effect on 949 observations, returning to being somewhat statistically significant with a p-value of less than 0.05.

For the men in particular these results are convincing. While these results fail to control for a fitness effect by way of heart rate data, they mirror well the analysis done by Guinness et. al. Despite losing some of the estimate's explanatory power by eliminating the proxy for fitness, being able to compare these results directly with the Guinness et. al. results allows for a more interesting dialogue regarding how the shoes affect amateurs and professionals differently. Focusing just on the men in both sets of analysis, it appears that amateurs actually experience a greater effect from the shoes. The original Guinness et. al. analysis estimated a 2.09% effect and the improvement estimated a 2.18% effect; this analysis, using an identical empirical model, estimates a 3.06% effect, a jump of at least 0.88%, if not closer to 1%. This difference in effect between elites and competitive amateurs raises plenty of fascinating hypotheses. It is possible that the shoes make a greater difference for amateurs because amateurs are naturally less efficient than professionals, resulting in greater efficiency gains from the shoes' mechanisms. Alternatively, the difference in effect could be something like a placebo effect for amateurs: perhaps amateurs believe that the effect exists more than professionals do. These suggestions are only hypotheses, though, and would require further analysis in order to prove.

Implications

While many of these results have their statistical and comparative significances, the practical significance surrounding the estimates is drawn from real world implications. The implications are manifold; these results carry one meaning for athletes, another for the shoe industry, and an entirely different meaning for econometricians.

First and foremost, these results are game changing for athletes, for better in some instances, and for worse in others. For amateur athletes, these results should be exciting. Depending on a male amateur's pace during his marathon, this improvement of 3.06% translates to a sizable time jump; for example, if an athlete believes he can run 6:00/mile without the shoes, this suggests to him that the shoes would enable him to run 5:50/mile with the Vaporflys on. That represents a jump from a 2:37:19 marathon to a 2:32:57 marathon, a difference of over four minutes. For the competitive amateur, that personal best becomes even more impressive. A less trivial example comes for the average male runner in the Strava dataset. That runner runs around 3:02 for the marathon, narrowly missing the qualifying window for the Boston Marathon, a major milestone for many runners. By putting on the Vaporflys, that runner could expect to run 2:56 instead, safely securing him a spot at Boston. For many, qualifying for Boston is a major achievement, and these shoes could give some runners the edge they need to make that dream a reality.

For the elites, though, the Vaporfly effect has mixed implications. On one hand, this technology is clearly pushing the sport forward. A marathon was run in under 2 hours for the first time ever in these shoes, breaking a barrier previously thought of as unbreakable. Despite marathon times at or below 2:05 having been run many times since Khalid Khannouchi set the world record in London in 2002, some professionals, and even scientists, had hypothesized that 2 hours was the magical barrier that humans would never cross. In October 2019, Eliud Kipchoge broke that barrier wearing the Vaporflys, reminding us through the process that "no human is limited." Additionally, professionals, just like amateurs, would all like to boast of fast personal records as possible, even if advanced technology is necessary to achieve it. On the other hand, however, comes the question of unfairness in the sport, a hot topic of debate within the professional running world at the moment. While pushing the sport forward is great, leaving athletes behind who cannot access the technology is not perceived as fair. Many professionals are restricted in their shoe choices by the shoe company they sign contracts with, so non-Nike runners are left at a disadvantage as long as the shoes are present. With the advantage evidently present and sizable, this technology seems to complicate the elite competition.

While Nike's shoes create this advantage at present, this advantage only exists so long as other shoe companies do not release equally, or greater, as effective of shoes. Thus, while there are questions surrounding the fairness of the Vaporfly, the effectiveness of the shoes also incites competition from other brands of running shoes. Other companies, such as New Balance and Hoka One One, have subsequently come out with their own 'super shoes' that are designed similarly in order to compete with Nike. In a sense, Nike has pushed the entire industry forward, with other companies following the path that they have created. The hope is that this competition between brands will ultimately eliminate any unfairness in the sport and continue to push the shoe technology forward. It is paramount that, if the sport is to be pushed forward from a technological perspective, that it be done equitably. We would rather see competition between individuals at the World Marathon Majors and the Olympic Marathon than see competition between two brands and their technology.

Finally, these results have minor implications for econometricians as well. As the research currently stands, this work is a case study in the difficulty of handling omitted variable bias. While attempts are made through both fixed effects and the fitness proxy variable, *z1suffer*, to control for the omitted variable of an athlete's fitness level, these attempts are not perfect, nor entirely successful. These attempts leave any good economist wanting a more compelling argument for how they effectively control for the bias. This difficulty also demonstrates a challenge in doing retrospective research this way: extensive data collection and analysis can be done, but if omitted variable bias is present, it might still taint the results. While there are still opportunities to try to handle the bias, the clear best route would be to run a randomized control trial on the shoes

Conclusions

Nike might be overconfident in their promise, but the data indicates that the Vaporflys will make you run faster. The results found in this study are strongly suggestive of a Vaporfly effect both existing and being practically significant. While confirming the magnitude of the effect with certainty is tempting, as the effect repeatedly appears in the data after rounds of analysis, the remaining questions regarding sources of bias make this claim somewhat dubious. Regardless, though, the results here suggest that Nike is overselling their promise that these shoes will make you 4% faster. This tagline is perhaps most relevant to competitive amateur male runners, as they saw the strongest effect, estimated at 3.1%, while all other types of runners (elite men and women, amateur women) in the dataset saw effects ranging from 1.3% to 2.2%. So, while the shoes will certainly help a runner set a personal best and run faster than they might otherwise be able to, Nike should revisit their claim.

This work, and all work studying the Vaporflys, is not yet complete. There remain plenty of areas for improvement in both the general study of the shoes and in the approach taken here. The best way to determine whether or not the effect really exists is to recruit people to a randomized control trial. Such a study might look something like this: recruit a number of people to run two marathons in a relatively short time window and randomly assign half of them the Vaporflys in the second marathon while the other half uses the same model of shoe they wore in the first half. The shoes worn in the first half of the study, as well as the second half for the control group, might look and vaguely feel like Vaporflys, perhaps just lacking the carbon fiber plate and the high-tech foam.

While there are plenty of areas for complications in that setup, such as athlete attrition or injury, it seems achievable. Unfortunately, though, the only company that would likely have the funding to secure such a study would be Nike itself, and Nike has no incentive to prove that their shoes make athletes any less than 4% faster.

With a randomized control trial off the table, retroactive studies like this one and the study put forth by Guinness et. al. are the best options for studying the Vaporfly effect at present. These studies could be improved upon in a multitude of ways.

Area for Improvement #1: More Data

Focusing on this study, one simple improvement would be to collect more data from the races that are being studied, prior to even expanding to other races. This study limited results collection to the top 2,000 times in each race reported on Strava's race pages, a method that was deficient in a at least two ways: first, this method heavily biased results towards men, leaving us with few women to study, and second, this approach resulted in fundamentally different cross-sections of the runners in each subsequent race, as athletes' posting of their results on Strava has been increasing over time. Removing that results limit would not only create a more representative cross-sample of the results, but would also increase the overall sample size, improving explanatory power. Starting with over 25,000 results, improving sample size did not seem like seem like a high priority until it became clear that only about one-third of athletes report the shoes that they wear, a piece of data that is clearly essential to the analysis.

Area for Improvement #2: Model Improvements

In addition to data collection improvements, there are a variety of model improvements and variations to be made. The glaring omission of this study is the failure to adequately handle for omitted variable bias. There are a couple of methods that might help handle the bias. First, no attempt was made in this study to use instrumental variables to handle the omitted variable bias, despite instrumental variables being the classic econometric way of handling it. While I do not have suggestions for how to use instrumental variables in this context, they are worth considering for future work. Then, for using proxy variables as a substitute for the omitted variable of fitness level, using a recent marathon time for each runner as a measure of fitness could also suffice. While this measure might have its own flaws, this proxy for fitness would tell a more convincing story of how omitted variable bias was handled in analysis, resulting in more compelling estimates.

Areas for Improvement #3: Expanded Focus

Continuations of this study could also go on to tackle more specific questions about the Vaporfly effect. The study from the *New York Times* put forth a variety of interesting estimates that would be worthwhile. For example, focusing more specifically on the effect that the shoes had when an athlete switches from another pair of shoes would be an interesting advancement that is possible within the current dataset. Additionally, building on the model that estimated the likelihood of setting a personal best while wearing the shoes would perhaps be the most relevant model for most runners; many runners might be convinced of the existence of an effect if the shoes could promise a personal best. Further, their studies also involved comparisons between the Vaporflys and other shoes. Comparing the effect that the Vaporflys have specifically with the effect of the other 'super shoes' that emerge, such as the Hoka One One Carbon X and the New Balance Fuel Cell, would become the most meaningful comparisons to make. As the competition between these shoe companies continue, consumers will need to do cost-benefit analyses of each of the shoes, comparing each shoe by a weighted average of their performance effect and their price point. Being able to assign performance effects to each of the shoes would be crucial in consumers' selection of the shoe that is right for them.

Even still, the Vaporfly effect is real and measurable, as this study has shown. The results found in this study convinced me to buy a pair of the shoes. While I have only worn the shoes a handful of times, the Vaporflys have already enabled me to complete runs I previously thought impossible. In a training run, I set a massive half marathon personal best in a controlled 14-mile effort. When running in the shoes, particularly when running at a moderate to hard effort, I felt like I was wearing springs on my feet. The uphills felt shorter and less intense and the downhills did not damage my legs as much as they normally would. Then, in the 2021 Providence Marathon, I debuted at the distance with a 2:38 performance.

Author (Will Peters) in the 2021 Providence Marathon wearing Nike Alphafly NEXT%'s

Through a rolling course on a sunny and warm day, sub-optimal conditions for a marathon, I ran faster than I had hoped months ago at the start of the training cycle. When the going got difficult late in the race, the shoes seemingly did the running for me, keeping me from completely imploding despite hamstring cramps and dehydration. By putting the shoes on and testing them for myself, I became even more convinced that they produce the promised effect. However, I am also currently in the best shape of my life without the shoes on, having also done some training runs in other shoes that were previously unimaginable. While I would like to attribute my breakthrough performances to the shoes for the sake of this study, I would be remiss to not take some of the credit myself after all of the hard work that went into this training cycle.

Again, we are left wondering how much of the effect can be attributed to the shoes, and how much comes from somewhere else. Has Nike innovated in a way that has truly changed the landscape of running, or are we simply misattributing advancements in the sport to technology instead of improved athletic ability? The short, unsatisfying, answer is that we are not yet sure; it seems like it might be a little bit of both. Even with the empirical results presented in this study, skepticism of the shoes is still valid. Either way, though, it is an exciting time to be an athlete, spectator, economist or consumer in this space. With or without the shoes, let barriers be broken, innovation be made and progress be pursued in the running world.

Appendix

Full list of marathons used in Guinness et. al. analysis and replication:

- Women's US Olympic Trials Marathon 2020
- Men's US Olympic Trials Marathon 2020
- The Marathon Project 2020 (*replication only*)
- Grandma's Marathon (2015-2019)
- New York City Marathon (2015-2019)
- Chicago Marathon (2015-2019)
- Houston Marathon (2015-2019)
- Cal Intl Marathon (2015-2019)
- Boston Marathon (2015-2019)
- Twin Cities Marathon (2015-2019)
- Philadelphia Marathon (2015-2019)
- Indianapolis Mon Marathon (2015-2019)
- Toronto Waterfront Marathon (2015-2019)
- LA Marathon (2015-2019)
- Richmond Marathon (2015-2019)
- Eugene Marathon (2015-2019)
- Phoenix Marathon (2015-2019)
- Marine Corps Marathon (2015-2019)
- Vancouver Intl Marathon (2015-2019)
- Ottawa Marathon (2015-2019)
- Columbus Marathon (2015-2019)
- Lakefront Marathon (2015-2019)
- Wineglass Marathon (2015-2019)
- Vermont City Marathon (2015-2019)

Bibliography

Quealy, Kevin, & Katz, Josh. "Nike Says Its $250 Running Shoes Will Make You Run Much Faster. What if That's Actually True?" *The New York Times* [New York City], 18 July 2018. *The New York Times*, Accessed 10 Dec. 2020.

Guinness, Joseph, et. al.. "An Observational Study of the Effect of Nike Vaporfly Shoes on Marathon Performance." *Cornell University*, 2 Oct. 2020, arxiv.org/abs/2002.06105. Accessed 10 Dec. 2020.

Hoogkamer, W., Kipp, S., Frank, J.H. *et. al.*. "A Comparison of the Energetic Cost of Running in Marathon Racing Shoes". *Sports Med* 48, 1009–1019 (2018). https://doi.org/10.1007/s40279-017-0811-2

O'Grady, T. J., & Gracey, D. (2020). "An Evaluation of the Decision by World Athletics on Whether or Not to Ban the Nike Vapor Fly Racing Shoe in 2020." *The Entertainment and Sports Law Journal*, *18*(1), 3. DOI: http://doi.org/10.16997/eslj.257

Dyer, B. "A Pragmatic Approach to Resolving Technological Unfairness: the Case of Nike's Vaporfly and Alphafly Running Footwear." *Sports Med - Open* 6, 21 (2020). https://doi.org/10.1186/s40798-020-00250-1

"Nike Zoom Vaporfly 4%." Nike News, Nike, 17 July 2017, news.nike.com/footwear/nike-zoom-vaporfly-4-release-date.